Taller evaluado 01-J 2026-1

CÓDIGO DE HONOR

"Como miembro de la comunidad de la Pontificia Universidad Católica de Chile, me comprometo a respetar los principios y normativas que la rigen. Asimismo, me comprometo a actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, al aprendizaje y la creación, difusión y transferencia del conocimiento."

INDICACIONES

Carácter: INDIVIDUAL.
Tiempo: 40 minutos.
Copiar y pegar el resultado desde la consola de R (las respuestas en Canvas tienen tolerancia al redondeo).
Está prohibido el uso de formularios y de archivos físicos personales. Solo está permitido estar en la interfaz de R (PositCloud o RStudio) o archivos de Canvas.

ENUNCIADO

Se cuenta con la base de datos de la Encuesta Nacional de Salud (ENS), una herramienta que utiliza el Ministerio de Salud para conocer qué enfermedades y qué tratamientos reciben hombres y mujeres de 15 años y más que viven en Chile. La información que arroja esta encuesta es de vital importancia para formular los planes de prevención, atención y políticas de salud para las personas que las necesitan.

La base de datos está en el archivo ENS.xlsx, que contiene la siguiente información (también puede usar estas bases que tienen otro formato pero la misma información: ENS.txt y ENS.rdata):

EDAD: Edad en años
NEDU: Años de estudios 1: <8 años, 2: 8-12 años, 3: >12 años
ZONA: Zona 1: Urbano, 2: Rural
REGION: 1: I. Tarapacá, 2: II. Antofagasta, 3: III. Atacama, 4: IV. Coquimbo, 5: V. Valparaíso, 6: VI. L. Bdo. O'Higgins, 7: VII. Maule, 8: VIII. Bíobío, 9: IX. La Araucanía, 10: X. Los Lagos, 11: XI. Aysén, 12: XII. Magallanes y Antártica, 13: XIII. Metropolitana, 14: XIV. Los Ríos, 15: XV. Arica y Parinacota
SEXO: Sexo 1: Hombre, 2: Mujer

GLUBASAL: Nivel de azúcar mg/dL
COLES: Colesterol Total mg/dL
HDL: Nivel de colesterol HDL mg/dL
LDL: Nivel de colesterol LDL mg/dL
TGD: Nivel de triglicéridos mg/dL
PAD: Presión arterial diastólica
PAS: Presión arterial sistólica
HTA: Hipertenso 1: Si, 0: No
SM: Sindrome Metabolico 1: Si, 0: No
FUMADOR: Fumador actual 1: Si, 0: No
DIABETES: Diabetico 1: Si, 0: No
PESO: Peso en kilos
TALLA: Talla en cm
CINTURA: Circunferencia de cintura en cm
CUELLO: Circunferencia de cuello en cm

Pregunta 1

Considere la variable **TALLA** y cree una nueva variable categórica llamada "**tallai**", definida de la siguiente manera:

- **baja**: si TALLA <= 154
- **media**: si 154 <TALLA <167
- **alta**: si TALLA >167

Luego, filtre únicamente los registros de las regiones: **1, 3, 5, 8, 12 y 13**, que corresponden a Tarapacá, Atacama, Valparaíso, Bío Bío, Magallanes y Antártica y Metropolitana, respectivamente.

A continuación, agrupe los datos según la condición de fumador actual (**FUMADOR**) y la categoría de talla (**tallai**). Finalmente, calcule para la variable de presión arterial sistólica **PAS** el coeficiente de variación (CV), con fórmula

$$CV = \frac{S_x}{\bar{X}}$$

Donde S_x es la desviación estándar muestral y $\bar{X}$ es la media muestral de la variable **X**.

Pregunta:
¿Cuál es el **menor coeficiente de variación** obtenido en los grupos formados?

Pregunta 2

Se desea estudiar la variable **X=PESO** mediante distintos modelos probabilísticos.
Suponga que **PESO** sigue una distribución **Log-Normal**:

$$X \sim \text{Log-Normal}(\texttt{meanlog = 4.4, sdlog = 0.2})$$

Pregunta:
Calcule el **rango intercuartílico** de la distribución, es decir,

$$IQR = Q_3 - Q_1$$

Donde

- Q_1 Q1 es el cuartil 1 o percentil 25
- Q_3 es el cuartil 3 o percentil 75.

Pregunta 3

Considere la variable **COLES** (colesterol) y evalúe visualmente qué conjunto de parámetros se ajusta mejor a los datos, comparando el histograma empírico con las curvas teóricas.
Si se asume que **COLES** sigue una distribución **Weibull** con los siguientes posibles parámetros: `shape` $\in \{4.6,\ 2.3\}$ y `scale` $\in \{212,\ 230\}$

Pregunta:
A partir de la superposición de curvas sobre el histograma de **COLES** con todas las posibles combinaciones de parámetros, indique qué valor del parámetro `shape` parece ajustarse mejor a los datos.

Pregunta 4

Considere nuevamente la variable **X=COLES**.
Suponga que **COLES** sigue una distribución **Weibull**

$$X \sim \text{Weibull}(\texttt{shape = } k, \texttt{scale = } 212)$$

Donde `shape`$= k$ es el mejor parámetro que ajusta la variable (esto se determinó en la pregunta anterior).

Pregunta:
Calcule la probabilidad de que un individuo seleccionado al azar tenga colesterol superior a 250, es decir:

$$P(X > 250)$$

Pregunta 5

Considere la variable **X=TGD** (nivel de triglicéridos, en mg/dL) y la variable **Y=ZONA** (1 = Urbano, 2 = Rural).

Considera dos grupos: el primero, **X** dado Y=1, es decir, el nivel de triglicéridos de las personas que viven en el área Urbana. El segundo, **X** dado **Y=2**, que serán los que viven en el área Rural.

Suponga que cada grupo puede modelarse mediante una distribución **Gamma**, cuyos parámetros se determinarán con las siguientes ecuaciones:

$$\texttt{shape} = \frac{\bar{X}^2}{S_X^2} \qquad\qquad \texttt{rate} = \frac{\bar{X}}{S_X^2}$$

Donde $\bar{X}$ es la media muestral y S_X^2 es la desviación estándar muestral de la variable **X**.
Pregunta:
Calcule los parámetros de la distribución Gamma para cada grupo y calcule el cociente de probabilidades.

$$\frac{P(X \leq 120|Y = 1)}{P(X \leq 120|Y = 2)}$$