Taller evaluado 01-M 2025-2

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.
Si el resultado numérico genera 4 o más decimales, aproximar hasta el cuarto decimal.
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

Parte 1

Las siguientes 3 preguntas, **que resultan en un puntaje total de 1pts**, respecta al cálculo de probabilidades. Para la información del uso de distribuciones en R, debe utilizar el comando "`? distributions`".

Pregunta 1 (0.25pts)

Si $\mathbf{X}$ proviene de una distribución Weibull con parámetro de forma (`shape`) igual a 10 y de escala (`scale`) igual a 7, calcule $P(X < 6)$

Pregunta 2 (0.25pts)

Si $\mathbf{X}$ proviene de una distribución Geométrica con parámetro de probabilidad (`prob`) igual a 0.6, calcule $\mathbf{P(X > 4 \,|\, X < 3)}$

Pregunta 3 (0.5pts)

Si $\mathbf{X}$ proviene de una distribución Normal con media (`mean`) igual a 3 y desviación estándar (`sd`) de 1, calcule $P(X < q \,|\, X \leq 3)$, donde q denota el cuantil 0.35 de $\mathbf{X}$

Parte 2

Las siguientes 8 preguntas, **que representan un puntaje total de 3pts**, buscan evaluar la importación y análisis de base de datos.

Un gimnasio quiere caracterizar a sus miembros para tomar decisiones con respecto a la infraestructura del mismo. Para ello, un día determinado, solicita a las personas que asisten al gimnasio información respecto a su entrenamiento, además de brindarles aparatos de monitoreo corporal (similar a relojes deportivos). La base de datos `Gym.csv` contiene dicha información. En particular, algunas de las variables se describen en la siguiente tabla.

Nombre Variable	Descripción	Valores
Age	Edad de la persona (en años)	Enteros positivos
Gender	Sexo de la persona	Male, Female
Avg_BPM	Latidos por minuto promedio de la persona	Enteros positivos
Calories_Burned	Calorías quemadas	Reales positivos
Experience_Level	Nivel de experiencia según su área	1, 2 o 3
Workout_Type	Tipo de entrenamiento	Strength, Cardio, HIIT, Yoga.
BMI	Índice de masa corporal	Reales positivos

0.1. Pregunta 4 (0.4 pts)

De los participantes, ¿cuántos de ellos tienen 40 años o más ("*Age*")?

Pregunta 5 (0.3 pts)

Con respecto a las calorías quemadas ("*Calories_Burned*"), calcule el promedio.

Pregunta 6

(0.3 pts) Con respecto a las calorías quemadas ("*Calories_Burned*"), calcule la desviación estándar.

Pregunta 7 (0.5 pts)

Suponga que la variable anterior ("*Calories_Burned*") proviene de una distribución **Gamma** con parámetro de forma (`shape`) igual a 11.03 y parámetro de tasa (`rate`) igual a 0.0122. Calcule el **cuantil** 0.9.

Pregunta 8 (0.5 pts)

Determine cuál es el nivel de experiencia ("*Experience_Level*") más común en la muestra. ¿Cuántas personas poseen dicho nivel?

Pregunta 9 (0.4 pts)

Calcule el promedio muestral de calorías quemadas ("*Calories_Burned*") para individuos con nivel de experiencia ("Experience_Level") igual a 1.

Pregunta 10 (0.3 pts)

Calcule el promedio muestral de calorías quemadas ("Calories_Burned") para individuos con nivel de experiencia ("*Experience_Level*") igual a 2.

Pregunta 11 (0.3 pts)

Calcule el promedio muestral de calorías quemadas ("Calories_Burned") para individuos con nivel de experiencia ("*Experience_Level*") igual a 3.

Parte 3

El archivo datos2.txt contiene la información de una sola variable ("variable") generada artificialmente de cierta distribución de probabilidad. En esta pregunta usted debe hallar los parámetros verdaderos con los que se obtuvo dicha muestra artificial.

Pregunta 12 (2 pts)

La base de datos `datos2.txt` consta de una sola variable ("*variable*"), la cual fue generada a través de una distribución Normal con media (`mean`) μ y desviación estándar (`sd`) σ. Si bien μ podría ser cualquier valor real y σ cualquier valor real positivo, usted debe asumir que μ y σ son enteros. Graficando curvas sobre el histograma de los datos, determine los valores de μ y σ utilizados para la generación de los datos. Luego, entregue el valor de σ hallado.

Hint: Para mayor precisión, puede utilizar el argumento `breaks` de la función hist().