Taller evaluado 01-W 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 9 preguntas, **que representan un puntaje total de 3.5pts**, 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 información respecto al entrenamiento de las personas que asisten al gimnasio, 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 minutos promedio de la persona.	Enteros postivos.
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.

Pregunta 1 (0.3pts)

Calcule la desviación estándar muestral de calorías quemadas (*Calories_Burned*) para personas con nivel de experiencia 1 (*Experience_Level==1*).

Pregunta 2 (0.3pts)

Calcule la desviación estándar muestral de calorías quemadas (*Calories_Burned*) para personas con nivel de experiencia 2 (*Experience_Level==2*).

Pregunta 3 (0.2pts)

Calcule la desviación estándar muestral de calorías quemadas (*Calories_Burned*) para personas con nivel de experiencia 3 (*Experience_Level==3*).

Pregunta 4 (0.5pts)

Sea $\bar{x}_1$ y s_1^2 el promedio y varianza muestral, respectivamente, de la variable *Calories_Burned* para individuos cuyo nivel de experiencia es igual a 1 (*Experience_Level==1*). Suponga que dicha variable (*Calories_Burned*) dentro del nivel de experiencia 1 proviene de una distribución Gamma con parámetro de forma (*shape*) igual a $\frac{\bar{x}_1^2}{s_1^2}$ y parámetro de tasa (*rate*) igual a $\frac{\bar{x}_1}{s_1^2}$. Calcule el cuantil 0.5.

Indicación: Calcule el promedio muestral mediante la función *mean()*, puesto que *summary()* pierde precisión.

Pregunta 5 (0.5pts)

Sea $\bar{x}_2$ y s_2^2 el promedio y varianza muestral, respectivamente, de la variable *Calories_Burned* para individuos cuyo nivel de experiencia es igual a 2 (*Experience_Level==2*). Suponga que dicha variable (*Calories_Burned*) dentro del nivel de experiencia 2 proviene de una distribución Gamma con parámetro de forma (*shape*) igual a $\frac{\bar{x}_2^2}{s_2^2}$ y parámetro de tasa (*rate*) igual a $\frac{\bar{x}_2}{s_2^2}$. Calcule el cuantil 0.5.

Indicación: Calcule el promedio muestral mediante la función *mean()*, puesto que *summary()* pierde precisión.

Pregunta 6 (0.5pts)

Sea $\bar{x}_3$ y s_3^2 el promedio y varianza muestral, respectivamente, de la variable *Calories_Burned* para individuos cuyo nivel de experiencia es igual a 3 (*Experience_Level==3*). Suponga que

dicha variable (*Calories_Burned*) dentro del nivel de experiencia 3 proviene de una distribución Gamma con parámetro de forma (*shape*) igual a $\frac{\bar{x}_3^2}{s_3^2}$ y parámetro de tasa (*rate*) igual a $\frac{\bar{x}_3}{s_3^2}$. Calcule el cuantil 0.5.

Indicación: Calcule el promedio muestral mediante la función *mean()*, puesto que *summary()* pierde precisión.

Pregunta 7 (0.4pts)

Utilizando los incisos anteriores, esto es, la distribución Gamma de la variable *Calories_Burned* dentro de cada nivel de experiencia (1, 2 y 3), y denotando X como las calorías quemadas e Y como el nivel de experiencia, calcule $\frac{P(X < 1000 \mid Y = 1)}{P(X < 1000 \mid Y = 2)}$.

Pregunta 8 (0.4pts)

Utilizando los incisos anteriores, esto es, la distribución Gamma de la variable *Calories_Burned* dentro de cada nivel de experiencia, y denotando X como las calorías quemadas e Y como el nivel de experiencia, calcule $\frac{P(X < 1000 \mid Y = 2)}{P(X < 1000 \mid Y = 3)}$.

Pregunta 9 (0.4pts)

Utilizando los incisos anteriores, esto es, la distribución Gamma de la variable *Calories_Burned* dentro de cada nivel de experiencia, y denotando X como las calorías quemadas e Y como el nivel de experiencia, calcule $\frac{P(X < 1000 \mid Y = 1)}{P(X < 1000 \mid Y = 3)}$.

Parte 2:

La base de datos **datos3.txt** consta de una sola variable ("*variable*"), la cual fue generada a través de una distribución Beta con primer parámetro de forma (*shape1*) α y segundo parámetro de forma (*shape2*) β. Si bien α y β podrían ser cualquier valor real positivo, usted debe asumir que $\alpha \in \{1, 2, 3\}$, mientras que $\beta \in \{1, 2, 3\}$. Graficando curvas sobre el histograma de los datos, determine los valores de α y β utilizados para la generación de los datos. Luego, responda las siguientes preguntas que representan un total de **2.5 pts**.

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

Pregunta 10 (0.8pts)

¿Cuál es el valor real de α que produjo la distribución de la única *variable* de la base de datos?

Pregunta 11 (0.8pts)

¿Cuál es el valor real de β que produjo la distribución de la única *variable* de la base de datos?

Pregunta 12 (0.9pts)

Evalúe la densidad de la distribución hallada en el punto 0.5. Reporte el resultado.