Taller Evaluado 01-L 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 personales. Solo está permitido estar en la interfaz de R (PositCloud o RStudio) y en archivos de Canvas.

ENUNCIADO

Parte 1

Las siguientes 4 preguntas, **que resultan en un puntaje total de 1 pts**, respectan al cálculo de probabilidades.
Para la información sobre el uso de distribuciones en R, pueden utilizar el comando **''**`? distributions`**''** para tener el listado de distribuciones disponibles en R Base.

Pregunta 1 (0.25 puntos)

Si $\mathbf{X}$ proviene de una distribución Geométrica con parámetro de probabilidad (`prob`) igual a 0.2, calcule $\mathbf{P(X < 9 \mid X > 7)}$.

Pregunta 2 (0.25 puntos)

Si $\mathbf{X}$ proviene de una distribución Binomial negativa con parámetro de tamaño (`size`) igual a 3 y de probabilidad (`prob`) igual a 0.5, calcule $\mathbf{P(X > 6)}$.

Pregunta 3 (0.25 puntos)

Si $\mathbf{X}$ proviene de una distribución Binomial con parámetro de tamaño (`size`) igual a 20 y probabilidad (`prob`) igual a 0.4, calcule $\mathbf{P(X = 8)}$.

Pregunta 4 (0.25 puntos)

Si $\mathbf{X}$ proviene de una distribución Poisson con parámetro de media (`lambda`) igual a 3, calcule $\mathbf{P(X = 5 \mid X > 2)}$.

Parte 2

Las siguientes 6 preguntas, **que representan un puntaje total de 3 pts**, buscan evaluar la importación y análisis de base de datos.

La base de datos `Coffee_sales.csv` contiene información sobre las ventas de cierta cafetería. Entre sus variables podemos encontrar la hora del día en que se realiza la venta, el tipo de café vendido, el día de la semana, etc. Específicamente, las variables medidas son las siguientes:

Nombre Variable	**Descripción**	**Valores**
hour_of_day	Hora del día en que se realizó la venta.	Enteros entre 1 a 24.
cash_type	Tipo de pago.	Único valor: card.
money	Cantidad pagada por café.	Reales positivos.
coffee_name	Tipo de café.	Americano, Latte, Cortado, etc.
Time_of_Day	Momento del día.	Afternoon, Morning, Night.
Weekday	Día de la semana.	Mon, Tue, Wed, Thu, Fri, Sat, Sun.
Month_name	Mes del año.	Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec.
Weekdaysort	Día de la semana (numérico).	Entero entre 1 a 7.
Monthsort	Mes del año (numérico).	Entero entre 1 a 12.
Date	Fecha en que se realiza la venta.	Fecha.
Time	Hora en que se realiza la venta.	Hora.

Concretamente nos enfocaremos en las variables *"hour_of_day"*, *"money"*, *"coffee_name"*, *"Time_of_Day"*, *"Weekday"* y *"Month_name"*, por lo que puede ser conveniente crear una base de datos solo con estas columnas.

Pregunta 5 (0.5 puntos)

¿Cuántas ventas se producen en el mes (*Month_name*) de mayor ventas?

Pregunta 6 (0.5 puntos)

Calcule el promedio de la variable *"money"*, interpretándose como la ganancia promedio por venta de café. Realice esto mediante el comando `mean()` de R para obtener el resultado exacto.

Pregunta 7 (0.5 puntos)

¿Cuál es la hora (*"hour_of_day"*) en la que se vende más café? El resultado debe ser un número entre 1 a 24.

Pregunta 8 (0.5 puntos)

Con respecto al tipo de café (*"coffee_name"*), determine cuál es el café que presenta más ventas. Ahora, ¿cuántas ventas le corresponden al café más consumido?

Pregunta 9 (0.5 puntos)

¿Cuánta ganancia total generan las ventas de café *"Cocoa"*?

Pregunta 10 (0.5 puntos)

Comparando con la mediana, ¿qué café (*"coffee_name"*) es el que tiene menor precio? (Americano, Americano with Milk, Cappuccino, Cocoa, Cortado, Espresso, Hot Chocolate, Latte)

Parte 3

El archivo `datos1.txt` contiene la información de una sola variable (*"variable"*) generada artificialmente de cierta distribución de probabilidad. En las siguientes preguntas usted deberá hallar los parámetros verdaderos con los que se obtuvo dicha muestra artificial.

Pregunta 11 (2 puntos)

La base de datos `datos1.txt` consta de una sola variable (*"variable"*) la cual fue generada a través de una distribución exponencial con parámetro de tasa λ (*rate*). Si bien λ podría ser cualquier valor real positivo, usted debe asumir que $\lambda \in \{\mathbf{0.5}, \mathbf{1}, \mathbf{1.5}, \mathbf{3}\}$. Graficando curvas sobre el histograma de los datos, determine el valor de λ utilizado para la generación de los datos.
Hint: Para mayor precisión, puede utilizar el argumento `breaks` de la función `hist()`.