Actividad 1 (FORO)

En esta actividad usted deberá aportar con 1 respuesta principal a la o las preguntas enunciadas por el profesor y comentar y/o intervenir 2 respuestas de sus compañeros (en otros hilos de conversación).
Además, junto a la respuesta principal, usted deberá subir un informe del trabajo realizado, el cual también **será considerado en la evaluación del foro**. Se permite 1 solo archivo adjunto (tamaño máximo 800MB).

I. Considere los datos 'hours_peer_week.csv', que contiene las horas que trabaja un grupo de trabajadores de EE.UU. a la semana.

1) Cargue el conjunto de datos en la sesión de trabajo de R usando la función `read.table`.
2) Calcule en forma manual el puntaje Z para las horas de trabajo semanal.
3) Construya un histograma de los datos originales y los datos estandarizados. Describa las características principales de los datos, comentando en el foro sobre la simetría y uni- o multi-modalidad de la distribución de los datos.
4) Construya un boxplot de los datos originales y los datos estandarizados. Comente sus resultados en el foro. ¿Existe evidencia de la presencia de "outliers"? Justifique su respuesta en el foro.
5) Repita los pasos 3) y 4) usando la función `scale`.

II. Considere los datos 'titanic.csv', sobre la tragedia del Titanic, cuya descripción se muestra en la siguiente tabla.

Variable	Descripción
passengerId	Identificador de pasajero
Survived	Variable que indica 1 si el pasajero sobrevivió y 0 si no.
Pclass	Clase del pasajero (1=primera clase, 2=segunda clase, 3=tercera clase)
Name	Nombre del pasajero
Sex	Género del pasajero
Age	Edad del pasajero
Sibsp	Número de hermanos o cónyuges a bordo
Parch	Número de padres o hermanos a bordo
Ticket	Número de ticket
Fare	Precio del ticket (en moneda local)
embarked	Puerto de embarque (C = Cherbourg; Q = Queenstown; S = Southampton)

Realice las siguientes actividades:

1) Cargue el conjunto de datos en la sesión de trabajo de R usando la función `read.table.`
2) Usando la función `summary(),` obtenga estadísticos descriptivos de las variables y discuta los resultados en el foro.
3) Cree una variable que indique el tamaño total de la familia del pasajero (incluyéndose él mismo).
4) Grafique la relación entre la tasa de sobrevivientes y el tamaño de la familia.
5) En base a lo observado en el punto anterior, proponga e implemente la discretización del tamaño de la familia. Justifique su decisión en el foro.
6) Identifique los pasajeros con datos faltantes para la variable `embarked` y `age` usando `la función is.na()`. ¿Qué tipo de mecanismo de generación de datos faltantes podría ser válido en cada caso? Justifique su respuesta en el foro.
7) Genere un conjunto de datos completos al imputar los valores faltantes de la variable `embarked` por la del puerto "C" y los valores faltantes de la variable edad por el promedio de edad de los datos observados.