Ejercicio – Imputación de datos

Este ejercicio consiste en analizar los valores perdidos de un conjunto de datos y utilizar las técnicas adecuadas de imputación aprendidas en clase para tratar dichos valores. Para esto, se utilizará una base de datos relacionada a las ventas (en miles de unidades) de un determinado producto en función del presupuesto de publicidad (en miles de dólares) invertido en televisión, radio y periódico. El archivo que contiene dicha información es **Advertising.csv**.

1. Cargar la base de datos y mostrar la estructura de las variables.
2. Convertir a lo mucho el 5% de valores de las variables Radio y TV a datos perdidos (NA). Guardar la nueva base de datos en un nueva data frame con el nombre **`publicidad`**.
3. Mostrar la proporción de datos perdidos por variable y por registro. Interpretar estos valores.
4. Analizar y visualizar el patrón de datos perdidos.
5. Realizar imputación simple usando la media y guardar la información en una nueva columna del dataset publicidad: **imp_mean**.
6. Realizar imputación por vecinos más cercanos empleando una cantidad adecuada de vecinos y guardar la información en publicidad (**imp_knn)**.
7. Comparar los datos imputados por la media y por vecinos más cercanos.
    a. ¿Qué tan diferentes son? Graficar para visualizar las observaciones imputadas por ambos métodos.
8. Se desean predecir las ventas en base a los valores invertidos en publicidad. Realizar una regresión lineal con los datos imputados (es su criterio elegir publicidad imp_mean o publicidad imp_knn).
9. Realizar imputación múltiple con el paquete **mice** y visualizar los datos imputados para cada variable y en cada muestra generada por la imputación. Comentar al respecto.
10. Elaborar una regresión lineal, similar a la del item 8, usando lo obtenido por imputación múltiple. Comparar los resultados con la regresión lineal anterior.