LABORATORIO V: Valores Faltantes (Preprocesamiento)

**INTRODUCCIÓN**

Esta práctica de laboratorio tiene como objetivo avanzar en la exploración de las técnicas de valores faltantes -análisis e imputación- de la etapa de Preprocesamiento, del Proceso de Descubrimiento de Conocimiento.

Para la exploración de estos temas, se utilizará el IDE R-Studio del lenguaje de programación R, a efectos de ejercitar los conceptos abordados en las clases teóricas.

**CONSIGNAS**

A partir del dataset *auto-mpg.data-original.txt[[1]](#footnote-1)*, se solicita trabajar sobre las siguientes consignas:

1. **SOBRE LOS DATOS**
   1. Cargue[[2]](#footnote-2) y explore el dataset: explique en qué consiste el mismo y qué características posee.
   2. Con las técnicas abordadas en las prácticas de laboratorio anteriores, realice un breve análisis exploratorio para identificar cual es la distribución de sus variables y si existe relación entre las mismas.
2. **VALORES FALTANTES**
   1. Verifique la existencia de datos faltantes en cada uno de los atributos ¿Existen datos faltantes en algún atributo? ¿En cuál/es? Indague sobre la proporción de datos que aparecen como faltante en la distribución.
   2. Aplique las técnicas de tratamiento de datos faltantes abordadas en clase (registros completos, sustitución por la media, e imputación por regresiones, hot deck[[3]](#footnote-3) y MICE).
   3. Analice gráficamente[[4]](#footnote-4) y analíticamente la variación en la distribución de datos de la variable estudiada. ¿Qué técnica de imputación afecta menos la distribución original?

Referencias sugeridas:

García, S., Luengo, J., & Herrera, F. (2016). Data preprocessing in data mining. Springer.

M. Brown, J.Kros (2003). Data mining and the impact of missing values.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

An Introduction to R: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>

1. Disponibles en: <https://archive.ics.uci.edu/ml/datasets/Auto+MPG> [↑](#footnote-ref-1)
2. Explore la instrucción *read.table().* [↑](#footnote-ref-2)
3. Algunas de las librerías disponibles para hot deck son *VIM (****recomendada****), HotDeckImputation y hot.deck.* [↑](#footnote-ref-3)
4. Se recomienda un gráfico de densidad, puede utilizar la instrucción *density*. [↑](#footnote-ref-4)