

PRÁCTICA – REGRESIONES MÚLTIPLES.

Librerías recomendadas: *dplyr*, *ggplot2*, *corrplot*, *ggcorrplot*, *e1071*, *GGally*, *tidyverse*, *ggpubr*, *base*, *car*, *MASS*, *leaps*, *hier.part*, *gvlma*, *lmtest*.

Considerar el dataset *state.x77* (en formato *data.frame*), que contiene información relativa a la esperanza de vida en cada uno de los 50 Estados de USA en función de variables tales como población, analfabetismo, ingresos, asesinatos, volumen de universitarios, días anuales con heladas, superficie del propio estado y densidad de población.

- 1.- Crear una nueva variable que indique la división a la que pertenece cada estado.
- 2.- Hacer un análisis de correlaciones (entre la esperanza de vida con el resto de atributos, y entre cada pareja de ellos), así como cualquier comentario apoyado en los parámetros estadísticos propios de un estudio descriptivo.
- 3.- Construir un modelo de regresión lineal múltiple (que relacione la esperanza de vida con el resto de las 8 variables). Comentar los resultados arrojados por el modelo en términos de R-cuadrado y mediante la evaluación de los correspondientes p-valores.
- 4.- Seguir un proceso step para seleccionar las variables más representativas en la predicción. Describir el modelo final resultante, hablando de los coeficientes obtenidos para las variables obtenidas como más relevantes, así como de la evolución del R-cuadrado ajustado y de los intervalos de confianza de los coeficientes.
- 5.- Diagnosticar el modelo final en términos de homocedasticidad y heterocedasticidad, normalidad de residuos y outliers en la predicción, aplicando cualquier contraste de hipótesis que se considere oportuno.