

Práctica de Estadística descriptiva

Ejercicio 1 (4 puntos)

La librería `ISLR` de R te permite cargar el set de datos `wage`. Responde a las siguientes cuestiones:

- a) Realiza un análisis descriptivo completo de los datos.
- b) Ajusta un modelo de regresión lineal para predecir ***logwage*** usando *age*, *jobclass*, *race* y *health_ins*.
- c) Interpreta los coeficientes del modelo y descríbelo en forma de ecuación. Calcula los intervalos de confianza de los coeficientes al 95%.
- d) ¿Cuáles de los predictores tienen una relación estadísticamente significativa con la variable respuesta? En base a esto, construye un modelo reducido que sólo emplee esos predictores, y compara su bondad de ajuste con el modelo anterior.
- e) Estudia la existencia de *outliers* o *high leverage points* en este último modelo.

Ejercicio 2 (4 puntos)

La librería `mlbench` de R te permite cargar el set de datos `BostonHousing`. Carga el *dataset* y sigue los siguientes pasos:

- a) Realiza un análisis descriptivo completo de los datos.
- b) Construye un modelo de regresión lineal que permita predecir *medv* (debes realizar el análisis completo: desde un modelo inicial hasta el ajuste del modelo final y la presentación de resultados, pasando por los diferentes pasos que te permitan optimizar tu primer modelo, así como comprobar si verifican o no las hipótesis en las que se fundamentan los modelos de regresión lineal).

Ejercicio 3 (2 puntos)

En la construcción del modelo del ejercicio anterior habrás pasado por varios modelos intermedios. Considera el modelo final y un par de esos modelos intermedios y, mediante *leave-one-out cross-validation* y *10-fold cross-validation*, estima su capacidad predictiva en situaciones reales y justifica si tu elección en el ejercicio anterior ha sido acertada o no.

La entrega de la práctica será individual. Debes entregar un fichero zip con tu nombre, por ejemplo, ***maria_del_mar_ruiz_martin_regresion.zip***.

Dentro de él, incluye un *pdf* explicando la resolución de los ejercicios, los ficheros de código R que utilices en la resolución, y referencias al código que uses para cada ejercicio. Si lo prefieres, también puedes hacer la práctica usando RMarkdown. En ese caso, entrega tanto el fichero *.Rmd* como el *.html* que se genera al hacer *knit*.