

## Entregable 1: Modelos de regresión lineal simple y múltiple

Ana Navarro Quiles

---

En la librería MASS puedes encontrar un famoso banco de datos llamado Cars93 que recoge información sobre 93 coches en venta en los Estados Unidos en 1993. La base contiene 27 variables relativas a 93 coches. Para saber qué información está contenida en las variables puedes escribir `?Cars93`. En este estudio vamos a usar la base de datos **`cars.csv`** que encontraréis en el aula virtual. Dicha base de datos contiene 15 de las 27 variables del banco de datos Cars93 indicado anteriormente.

---

Ejercicio 1: Considera la variable respuesta `Price` relacionandola con el predictor `MPG.city`.

1. Evalúa el efecto de `MPG.city` sobre `Price`.
2. Obtén la recta mínimos cuadrados. Interpreta los resultados obtenidos.
3. Dibuja el diagrama de dispersión, la recta de regresión y las bandas de confianza al 90 %.
4. ¿Te parece adecuado haber realizado regresión lineal o es preferible otro tipo de regresión?. Ajusta el modelo que te parezca más adecuado y, en dicho caso, dibuja nuevamente las bandas de confianza correspondientes al 90 %.
5. ¿Qué precio mínimo se espera para aquellos coches con un consumo de 12 litros a los 100 km por ciudad? Calcula e interpreta el intervalo de confianza y el de predicción.

Ejercicio 2:

1. Considerando un tope de 10 variables, encuentra el número óptimo de variables a incluir en un modelo predictivo de `Price`, según los criterios  $R^2$ , BIC y AIC.
  - ¿Qué variables incluye el modelo obtenido?. Interpreta los coeficientes obtenidos, ¿consideras que tiene sentido?.
2. Selecciona el mejor modelo con el método stepwise.
3. Selecciona el mejor modelo con el método stepwise considerando la variable `Passengers` como factor.

4. Depura el modelo anterior (apartado 3.) sólo si te parece oportuno y contesta a las siguientes preguntas:
    - ¿Qué % de la varianza de Price explica el modelo?
    - ¿Cuál es la variable menos explicativa?. ¿Cuál es su efecto sobre Price?
  5. Realiza el diagnóstico de tu modelo, sin emprender ninguna acción, e indica los problemas que presenta. *Ayuda: Para ver los valores influyentes podéis utilizar los comandos `influencePlot` e `influence.measures` de la librería `car`. Para la normalidad usar el `qqplot` y/o el test de Shapiro-Wilk. Para la linealidad y homocedasticidad evaluar los residuos.*
  6. Emprende ahora las acciones que te parezcan oportunas e indica los problemas que has conseguido solucionar o disminuir.
  7. Con el modelo final obtenido, obtén la predicción del precio para un coche en la mediana de los predictores.
-