

# Aprenentatge màquina (I). Máster en Ciencia de Datos. ETSE. Universitat de València

## Tarea: 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 la variable X con la que tenga mayor relación lineal.

1. Evalúa el efecto de X sobre Price.
2. Obtén la recta mínimos cuadrados. Interpreta los resultados obtenidos (coeficientes, significatividad,  $R^2$ , contraste del modelo, etc...).
3. Dibuja el diagrama de dispersión, la recta de regresión y las bandas de confianza al 90 %.
4. Realiza un diagnóstico de los residuos. Si falla algunas de las condiciones, busca una posible solución.

Ejercicio 2: 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 (coeficientes, significatividad,  $R^2$ , contraste del modelo, etc...).
3. Dibuja el diagrama de dispersión, la recta de regresión y las bandas de predicción al 90 %.
4. Realiza un análisis de los residuos.
5. ¿Te parece adecuado haber realizado regresión lineal o es preferible otro tipo de regresión?. Ajusta el modelo que te parezca más adecuado.
6. ¿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 3:

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? (seleccionar el criterio que más te guste). Interpreta los coeficientes obtenidos, ¿consideras que tienen 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. Contesta a las siguientes preguntas:
    - ¿Qué % de la varianza de Price explica el modelo?
    - ¿Podrías depurar el modelo?
    - ¿Cuál es el efecto de la variable Origin sobre Price?
  4. ¿Qué modelo de los apartados anteriores es mejor? Con el que te quedes, realiza el diagnóstico de tu modelo, sin emprender ninguna acción, e indica los problemas que presenta.
  5. Emprende ahora las acciones que te parezcan oportunas e indica los problemas que has conseguido solucionar o mejorar un poco.
  6. Obtén la predicción del precio para un coche en la mediana de los predictores en el modelo escogido. *Notar que las variables categóricas se tratan de diferente manera, no hay mediana.*
-