

Práctica 4: Regresión lineal múltiple

Módulo de Modelos Lineales.
Máster de Bioestadística, Universitat de València.

Miguel A. Martinez-Beneito

Tareas

1. Para el banco de datos **Auto** de la librería **ISLR**, considera ahora un modelo de regresión lineal múltiple sobre **mpg** empleando como covariables, sin interacción, el resto de variables numéricas del banco de datos. Interpreta los resultados obtenidos: ¿Qué variables tienen un efecto significativo sobre **mpg**? ¿De qué forma influyen? Respecto al efecto de **horsepower**, que ya fue estudiado en la práctica anterior ¿Cómo cambia la interpretación del efecto de dicha variable del análisis anterior a éste? ¿Cómo ha cambiado la variabilidad de su efecto?
2. Volviendo al modelo que hayas ajustado en la tarea anterior, considera ahora ese mismo modelo eliminando aquellas variables que no tuvieran efecto significativo sobre **mpg**.
 - ¿Cómo varía la estimación y variabilidad de dichos efectos al eliminar el resto de variables del modelo?
 - ¿A qué crees que se debe la distinta disminución en la varianza de los estimadores en este nuevo modelo?
 - ¿Consideras aconsejable considerar en el modelo la interacción de las variables que hubieras incluido en el modelo anterior? Interpreta el efecto de la interacción que hayas incluido sobre **mpg**.
3. El archivo **Nitrite.Rdata** contiene datos sobre el consumo de nitritos de distintas plantas de alubias en función de la intensidad de la luz.
 - Representa la nube de puntos correspondiente a este banco de datos, ignora la variable día para el análisis estadístico.
 - Ajusta la relación polinomial que consideres más adecuada para representar la relación entre el uso de nitritos e intensidad de la luz.
 - Representa gráficamente la relación polinomial que te haya parecido oportuna, así como la relación correspondiente al resto de polinomios de orden menor que hayas ajustado para observar las diferencias que se producen entre los distintos ajustes.
 - Para el modelo que consideres más adecuado, representa la relación ajustada entre nitritos e intensidad de luz, así como sus intervalos de confianza y predicción al 95%.
4. Plantéate la veracidad o falsedad de las siguientes afirmaciones:
 - La suma de cuadrados residual siempre disminuirá al incluir nuevas variables en un modelo de regresión lineal múltiple.
 - La variabilidad de los coeficientes de un modelo de regresión siempre disminuirán al incluir nuevas variables en el modelo.
 - La variabilidad de los coeficientes de un modelo de regresión siempre aumentarán al incluir nuevas variables en el modelo.
 - Si la correlación de la variable respuesta con cierta covariable es positiva, el correspondiente coeficiente en un modelo de regresión lineal múltiple será también positivo.