

Práctica 5: Validación de modelos de regresión

Módulo de modelos lineales.
Máster de Bioestadística, Universitat de València.

Miguel A. Martinez-Beneito

Tareas

1. Para el modelo lineal que ajustaste en la tarea 1 de la práctica 3 (relación entre `mpg` y `horsepower` para el banco de datos `Auto`) valora su ajuste, validando las hipótesis del modelo de regresión una vez ajustado el modelo. Repite dicha validación para el modelo de la tarea 3 en el que se asumía una relación lineal entre `mpg` y `1/horsepower`.
2. El banco de datos `Congenitas.Rdata` contiene las defunciones por enfermedades congénitas ocurridas en España desde el año 1980 hasta 2018, en ese mismo orden. Dado el alto número de defunciones observadas durante el periodo de estudio resulta razonable tratar dicha variable como continua y modelizarla mediante un modelo de regresión lineal Normal.
 - Ajusta un modelo de regresión lineal que modeliza las defunciones del banco de datos como función del año del periodo de estudio $x=2:39$. Considera un ajuste cuadrático de esta variable explicativa y representa el ajuste obtenido.
 - Evalúa la hipótesis de Normalidad para el modelo que acabas de ajustar.
 - Evalúa la hipótesis de homocedasticidad de los datos alrededor de la curva ajustada ¿a qué crees que se puede deber la heterocedasticidad de los datos?
 - Representa los residuos estandarizados y valora sobre dicha representación la adecuación de la hipótesis de linealidad de la tendencia conforme avanza el periodo de estudio.
 - Considera un ajuste polinómico de orden superior, hasta el grado que consideres oportuno. Representa el ajuste del nuevo modelo que hayas ajustado.
 - ¿Desaparecen los problemas de homoscedasticidad y tendencia de los residuos para este nuevo modelo?
3. El banco de datos `estimulos.Rdata` contiene dos variables, ambas son el tiempo de respuesta del cerebro ante distintos estímulos. La variable `x` es el tiempo de respuesta a un sonido acompañado de un estímulo visual y la variable `y` es el tiempo de respuesta únicamente al estímulo visual, para una serie de 50 individuos.
 - Representa gráficamente ambas variables.
 - Ajusta un modelo de regresión lineal para explicar la variable `y` en función de `x`. Comprueba que la relación funcional entre ambas variables es la adecuada y que no resulta necesario incluir un polinomio de mayor grado en función de `x`.
 - Evalúa la hipótesis de Normalidad para el modelo que acabas de ajustar.
 - Determina la transformación de Box-Cox que mejoraría en mayor medida la Normalidad de la variable respuesta `y`. Ajuste de nuevo un nuevo modelo de regresión lineal sobre la nueva variable transformada ¿Observas mejora en cuanto a la hipótesis de Normalidad que habías evaluado anteriormente?