

Práctica para Modelos Lineales

Modelo Lineal Múltiple Con Interacción

Una bandada de arrendajos azules (*Cyanocitta cristata*) ha decidido asentarse en el arboreto de la Escuela de Ingeniería Forestal y del Medio Natural. Como buenos ~~frikis~~ ^{frikis} apasionados de la naturaleza, hemos ido a tomar datos de esta especie norteamericana. Tras varias mañanas recogiendo datos, observamos que estos pájaros se alimentan de larvas escondidas en la corteza de los árboles, por lo que pensamos que puede haber una relación positiva entre la longitud del pico y la masa corporal de los arrendajos. Sin embargo, también hemos observado que las hembras pasan más tiempo comiendo, por lo que esperamos que la relación entre la longitud del pico y la masa del individuo sea distinta entre sexos (si todas las hembras comen más, la relación entre la longitud del pico y su masa será menos apreciable que en machos).



Habéis incluido vuestra base de datos en el paquete **Stat2Data**, para que todo el mundo pueda encontrarla con el nombre "BlueJays".

1. Comprende y explora la base de datos. ¿Qué tamaño muestral hemos conseguido?
2. Define las hipótesis que te permiten resolver nuestra duda.
3. Comprueba que las asunciones del análisis estadístico a realizar no son violadas.

1. **Normalidad de los residuos del modelo:** Aunque parezca contradictorio, para calcular los residuos de un modelo, es necesario ajustar el modelo previamente. Una vez tengas el modelo definido como una variable, utiliza la función `resid()`. Para comprobar su normalidad puedes utilizar un histograma o un qqplot.

2. **Outliers:** Para comprobar si existen valores atípicos en nuestra base de datos, es muy útil representar las distintas variables continuas mediante boxplots (también llamados diagrama de caja y bigotes). ¿Recuerdas qué representa cada uno de los elementos de un boxplot? ¿Tienes algún outlier en tu bd?

¡Prueba con la función `boxplot(y~x)`!

3. **Homocedasticidad:** Realmente hay una manera más sencilla de comprobar las asunciones de un modelo (jeje), ¡pero así ya has aprendido a cómo checkear la normalidad de los residuos y la presencia de outliers manualmente en los pasos anteriores!

La función `plot()` aplicada sobre un modelo nos grafica cuatro plots con los diagnósticos del modelo. Son algo difíciles de interpretar, pero vamos allá... El primer plot grafica los valores de los residuos contra los valores reales u observados, y nos permite comprobar la asunción de la homogeneidad de varianza. Para no violar esta asunción, este plot tiene que parecer una "noche estrellada" (no debe haber patrones en los puntos, y la línea roja debe ser más o menos horizontal, sin curvas demasiado acentuadas). El segundo plot es prácticamente lo mismo a lo que hemos hecho en el primer paso, un qqplot de los residuos del modelo. Es decir, nos permite comprobar la normalidad de los residuos. Los puntos tienen que quedar en línea recta sobre la diagonal. El tercer plot no nos interesa ahora mismo. Y finalmente, el cuarto plot nos muestra aquellas observaciones o datos que sean atípicos, es decir, outliers. En nuestro caso, no tenemos ninguno.

Si quieres leer un poco más sobre la interpretación y significado de los diagnósticos de modelos, puedes hacerlo aquí: <https://data.library.virginia.edu/diagnostic-plots/>

4. Desarrolla los análisis estadísticos que te permitan testar tus hipótesis y saber si existe una relación entre la longitud del pico de los arrendajos y su masa, y si esta relación depende del sexo. ¿Qué proporción de varianza de la masa de los arrendajos está explicada por tu modelo?

5. Crea un gráfico que te permita visualizar los resultados y llegar a una conclusión respecto a tus hipótesis.