

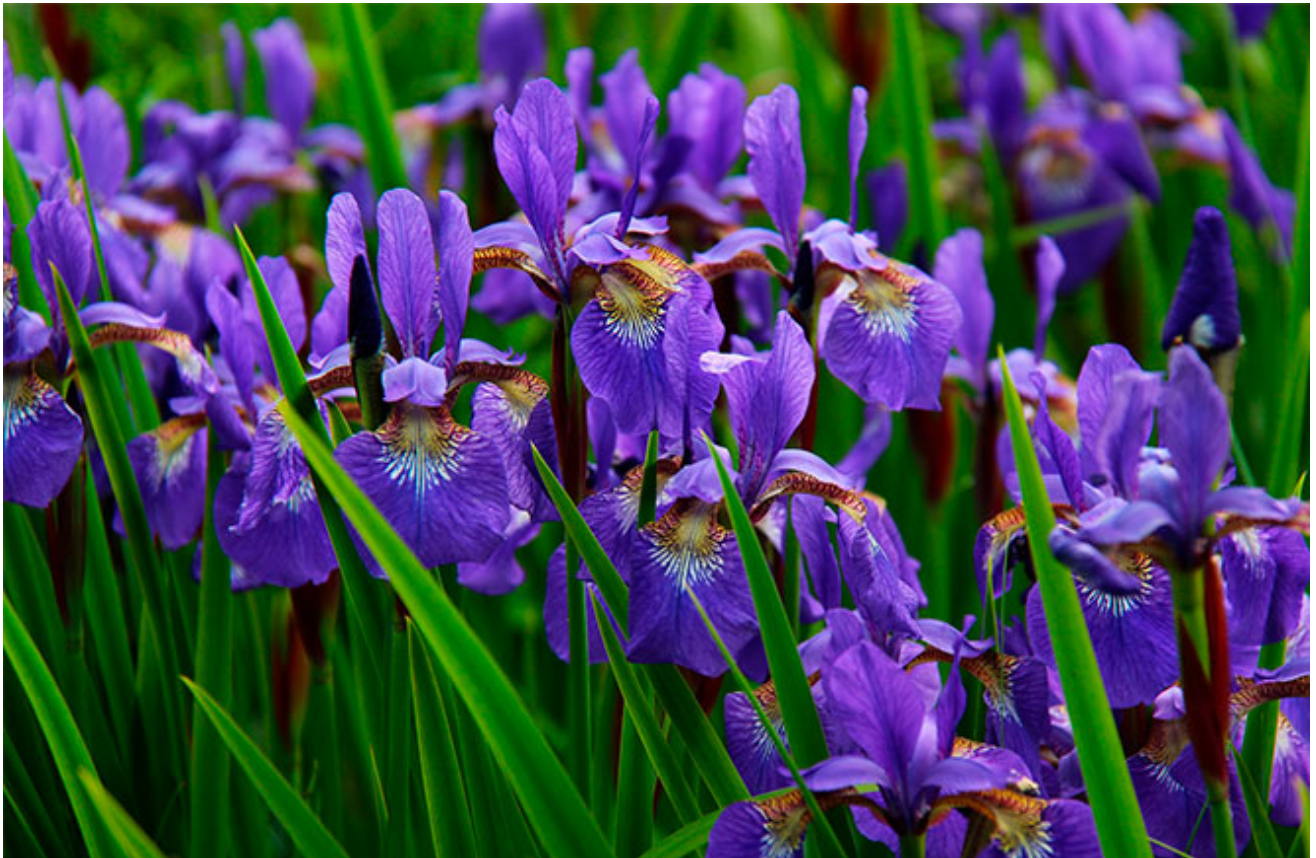
# Práctica para Modelos Lineales

---

## Modelo Lineal Simple

El perianto es la estructura floral que envuelve los órganos sexuales, es decir, la parte no reproductiva de la flor. Frecuentemente, el perianto está compuesto de sépalos (parte que conforma el cáliz y cubre el resto de partes de la flor hasta que se abre) y pétalos. Sin embargo, varios estudios (e.g. Irish, 2009- *Journal of Experimental Botany*) sugieren que no siempre fue así, ya que parece que las primeras angiospermas no tenían un perianto diferenciado. Por lo tanto, pese a que estas partes de la flor comprenden distintas funciones actualmente, cabría esperar que su morfología sea semejante, debido a compartir un origen evolutivo común.

Como aficionad@ a la botánica, llevas años recogiendo la longitud y anchura de los pétalos y sépalos de flores de iris, lo cual te permite testar esta hipótesis, respondiendo a la pregunta *¿existe una relación morfológica entre los sépalos y los pétalos?*



Como eres super generos@, has donado tu base de datos a R, y la puedes encontrar en el paquete **datasets**, bajo el nombre "iris".

1. Comprende y explora la base de datos. ¿Qué tamaño muestral has conseguido a lo largo de todos estos años?
2. Qué longitud media tienen los pétalos? ¿Qué anchura mínima y máxima tienen los sépalos?
3. Define las hipótesis que te permiten responder a la pregunta, tanto para la anchura, como para la longitud de los elementos florísticos.

4. Comprueba que las asunciones del análisis estadístico a realizar no son violadas. Utiliza métodos gráficos y estadísticos.

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 checkar 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.

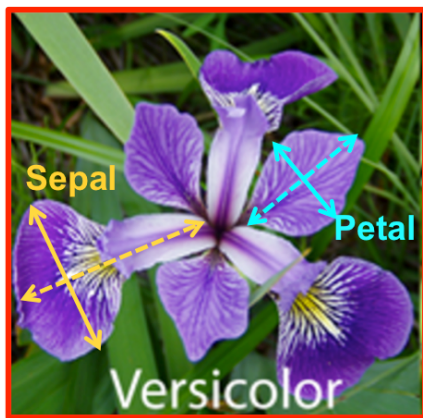
Si quieres leer un poco más sobre la interpretación y significado de los diagnósticos de modelos, puedes hacerlo aquí: <https://data.library.virginia.edu/diagnostic-plots/>

5. Desarrolla los análisis estadísticos que te permitan testar tus hipótesis y saber si existe una relación entre el tamaño de los sépalos y los pétalos de tus iris. ¿Cuál es la ecuación que define la relación entre la longitud de los pétalos y los sépalos? ¿Qué proporción de varianza de la anchura de los sépalos está explicada por tu modelo?
6. Crea un gráfico que te permita visualizar los resultados y llegar a una conclusión respecto a tus hipótesis.

---

## Modelo Lineal Múltiple con Interacción

Espera un momento... ¿quééééé? Te acabas de dar cuenta de que además de los valores morfológicos de los iris, ¡también habías identificado las especies que mediste!



Esto te permite desarrollar mucho más tu modelo, ¡que fantasía! Vamos a seguir donde lo dejamos, pero esta vez incluiremos un detalle a nuestra pregunta de investigación: ¿difiere la relación morfológica entre sépalos y pétalos entre las especies?

Como seguimos trabajando con el mismo dataset, no hace falta que comprobemos las asunciones otra vez, ¡pero recuerda hacerlo **siempre** que vayas a correr un nuevo modelo!

1. Define las hipótesis que te permiten contestar a la pregunta anterior.
2. Desarrolla el análisis estadístico que permite resolver tu pregunta e interpreta los resultados. ¿Cuáles son las ecuaciones obtenidas para las distintas especies de iris?
3. Con los anteriores resultados, ¿qué conclusión obtenemos en relación a nuestra hipótesis?
4. Crea un gráfico que permita visualizar los resultados diferenciando las observaciones (o datos) en función de la especie.