

# Técnicas estadísticas avanzadas para la conservación de la biodiversidad - Ejercicios

## Estadística descriptiva

- Considera los conjuntos de datos `nations.csv`, `iris.csv`, y `trees.csv`, en la carpeta `datasets`. Para estos conjuntos de datos,
  - selecciona una o varias variables numéricas.
  - calcula las medidas de centralidad y dispersión de las variables seleccionadas (media, mediana, moda, desviación típica, asimetría, curtosis).
  - extrae una muestra de cada variable, y calcula la media muestral, la desviación típica muestral, el error estándar de la media, y un intervalo de confianza al 95% para la media.

## Diseño experimental

- Considera un experimento en el que quieres medir el efecto que producen los fertilizantes disueltos en agua sobre las comunidades de invertebrados de un río. En concreto, quieres saber si el porcentaje del herbicida RoundUp (e.g. [https://en.wikipedia.org/wiki/Roundup\\_\(herbicide\)](https://en.wikipedia.org/wiki/Roundup_(herbicide))) disuelto en agua influye sobre la riqueza de especies de invertebrados en agua. Tienes disponibles 20 tanques en los que puedes replicar las comunidades bióticas del río, y alterar la composición química del agua a voluntad. También tienes disponible un técnico que se ocupa de recoger y analizar muestras durante un año. ¿Cómo diseñarías el experimento?

## Tests de hipótesis

- Considera el conjunto de datos `nations.csv`.
  - genera una gráfica con la distribución de producto interior bruto (variable `gdp`), diferenciando los grupos que creas conveniente (e.g. por años, o por regiones).
  - ¿sigue el producto interior bruto de los diferentes países una distribución normal?
  - ¿hay diferencias estadísticas entre el `gdp` de los países de la región europea y los de la región del sur de asia?
  - ¿está correlacionado el `gdp` con la esperanza media de vida (variable `life_expect`)?
- Considera el conjunto de datos `iris.csv`.
  - ¿qué variables están más y menos correlacionadas entre sí?
  - ¿tiene la especie *Iris setosa* una longitud de pétalo diferente a la especie *Iris versicolor*?
- Considera el conjunto de datos `earthquakes.csv`.
  - ¿hay diferencias entre la magnitud media de terremotos registrados en Eurasia y en América del Sur?

## Modelos lineales

- Considera el conjunto de datos `iris.csv`.
  - ¿se puede predecir la longitud de pétalo de una flor (petal length) a partir de su anchura (petal width)?
  - ¿la relación entre longitud y anchura de pétalo es igual o diferente para las tres especies de Iris?

- ¿cuál es la longitud media esperada de pétalos que tengan anchura 2.6? ¿y su intervalo de confianza?
- Considera el conjunto de datos `starwars_info_personajes.csv`.
  - ¿existe relación entre la altura (`height`) y el peso corporal (`mass`) de los personajes? ¿cómo de buena es la relación?
  - ¿hay algún problema o elemento a tener en cuenta en los datos?

## Modelos lineales generalizados

- Considera el conjunto de datos `titanic_long.csv`. Genera GLMs para responder las siguientes preguntas:
  - ¿cuál es la probabilidad de que un pasajero de primera clase sobreviviera al naufragio? ¿y la probabilidad para un pasajero de la tripulación (`crew`)?
  - ¿hubo diferencias en la probabilidad de supervivencia de hombres y mujeres?