



## Guía 7: Modelo lineal múltiple

Laboratorio de datos 2023 (comisión: G. Solovey)

Para resolver estos ejercicios, cargar la librería `palmerpenguins`, usar el dataset `penguins`, borrando las observaciones que tengan algún `NA`.

### Ejercicio 1

---

Se quiere predecir el peso de un pinguino usando como variables predictoras el largo de la aleta y el sexo del pinguino.

- Escribir (en lápiz y papel) la ecuación de un modelo lineal para este caso. ¿Qué unidades tienen las variables y cómo se codifica la variable “sexo del pinguino”?
- Ajustar el modelo usando todos los datos disponibles. Reportar los coeficientes encontrados y calcular el error de predicción.
- Hacer un gráfico de los datos junto con las predicciones del modelo.
- Dos pinguinos que tienen igual largo de aleta, uno macho y otro hembra, ¿qué diferencia de peso predice el modelo que tendrán?

### Ejercicio 2

---

Ahora se quiere predecir el peso de un pinguino usando como variables predictoras el largo de la aleta y la especie del pinguino.

- Escribir (en lápiz y papel) la ecuación de un modelo lineal para este caso. ¿Cómo se codifica la variable “especie”?
- Explicar qué diferencia tiene este modelo respecto al propuesto en el ejercicio 1.
- Ajustar el modelo usando todos los datos disponibles. Reportar los coeficientes encontrados y calcular el error de predicción.
- Hacer un gráfico de los datos junto con las predicciones del modelo.

### Ejercicio 3

---

Modificar lo hecho en el ejercicio 2 introduciendo una tercera variable predictora que sea la “interacción” entre el la aleta y la especie.

- Escribir (en lápiz y papel) la ecuación de un modelo lineal para este caso.
- Explicar qué diferencia tiene este modelo respecto al propuesto en el ejercicio 2.

- c. Ajustar el modelo usando todos los datos disponibles. Reportar los coeficientes encontrados y calcular el error de predicción.
- d. Hacer un gráfico de los datos junto con las predicciones del modelo.
- e. Dos pingüinos que tienen igual largo de aleta, uno Adelie y otro Gentoo, ¿qué diferencia de peso predice el modelo que tendrán?

## Ejercicio 4

---

Crear varios modelos de regresión lineal múltiple para predecir el peso de un pingüino en base a varias variables predictoras.

- a. Dividir el dataset en un grupo de entrenamiento y uno de test (80% - 20%).
- b. Crear y ajustar 6 modelos que contengan desde 1 hasta 6 variables predictoras sumando en cada modelo una variable nueva. Las 6 variables son: sexo, especie, año de la expedición, largo de la aleta, largo del pico, ancho del pico.
- c. Calcular para cada uno el error predicción en el grupo de entrenamiento y en el grupo de test.
- d. ¿Cuál modelo tiene el menor error de ajuste? ¿Cuál el menor error de predicción?