



## Tarea C02. Inferencia Estadística para análisis de datos

SciData Math, Septiembre 2022

En la sección de **Data** de nuestro repositorio encontrarás los tres archivos para esta tarea, de los cuales dos ya hemos trabajado. Los tres archivos son *cars.csv*, *online\_retail2.txt* e *iris.csv*. De esos tres, el que no hemos trabajado es el de iris.

Te pongo en contexto: iris es un tipo de flor la cual, entre otras, tiene tres especies (revisa la columna species de dicho dataframe). La tabla de iris fue formada tomando 50 flores de cada una de las especies; para cada una de las 150 flores se midió el largo del sépalo (*sepal\_length*), el ancho del sépalo (*sepal\_width*), el largo del pétalo (*petal\_length*) y el ancho del pétalo (*petal\_width*).

Una vez que tienes las tres bases de datos, realiza las siguientes gráficas:

1. De la tabla *cars*, realiza un gráfico de puntos con el precio y la potencia de cada auto (la potencia son los caballos de fuerza; es decir, la columna de horsepower).
2. De la tabla de iris, realiza el gráfico de densidad de la longitud del sépalo de las 150 flores.
3. De la tabla de iris, realiza un gráfico donde muestres las densidades la longitud del sépalo de las 150 flores separadas por especie.
4. Nota que en la tabla *online\_retail2* aparecen las fechas de transacciones en formato «día/mes/año hora». Grafica cómo se comportaron los

precios (columna *Price*) mes a mes durante 2011 para Reino Unido. Un posible camino para resolver este problema es:

- Crea una tabla con los datos de Reino Unido (es cuando la columna *Country* es igual a *United Kingdom*).
- De la columna *InvoiceDate*, utilizando el comando de **R** llamado *substr* extrae el mes y el año.
- Crea una nueva tabla con la información de 2011.
- El total de ventas mes a mes se puede calcular con la función *tapply* discriminando por mes.
- Continúa. Si tienes dudas, recuerda que puedes publicarlas en el foro.