



IIC2115 – Programación como Herramienta para la Ingeniería (II/2025)

Ejercicio Formativo 1 Capítulo 1

Aspectos generales

- **Objetivos:** familiarizarse los elementos básicos del análisis de datos tabulares.
- **Entrega:** lunes 25 de agosto a las 17:30 hrs. en repositorio privado y ticket de salida.
- **Formato de entrega:** archivo E1.ipynb con los solicitado, ubicado en la carpeta **C2** del repositorio.

Descripción del problema (versión normal)

En este ejercicio trabajará con el conjunto de datos *Anscombe's quartet*, compuesto por cuatro series de pares (x, y) con propiedades particulares. Su tarea consiste en explorar estadística y visualmente estos conjuntos, calculando métricas relevantes y comparándolas entre sí, para luego analizarlas críticamente y discutir qué revelan (y qué esconden) sobre los datos. Además, deberá complementar el análisis con visualizaciones que permitan caracterizar mejor las diferencias entre los conjuntos y experimentar con la eliminación e imputación parcial de datos, reflexionando sobre cómo esto afecta los resultados.

Descripción del problema (versión guiada)

El conjunto de datos Anscombe's quartet, disponible en el archivo `E1_data.csv`, consiste en 4 conjuntos de una serie de pares (x, y) , con algunas características particulares. En este ejercicio explorará algunas de ellas mediante el análisis estadístico y visual.

Siga las instrucciones de la lista de a continuación, recordando siempre hacer *commit* regularmente en el repositorio privado, usando comentarios descriptivos:

1. Cree una carpeta de nombre **C2** dentro del repositorio privado y agregue un nuevo **Jupyter Notebook** con nombre `E1.ipynb`.
2. Cree una celda de texto dentro del documento e ingrese el título de la actividad “Ejercicio Formativo 1 Capítulo 2” junto con tu nombre.
3. Cargue el conjunto de datos utilizando Pandas e imprima su contenido. Comente acerca del formato utilizado para los datos.
4. Calcule algunas métricas interesantes para obtener mayor conocimiento de estos datos. Es importante notar que se esperan los valores de las métricas por conjunto, es decir, debe obtener las métricas para el conjunto A, B, C y D de manera separada. Considere inicialmente las siguientes:
 - Media.
 - Desviación estándar.
 - Máximo.
 - Mínimo.
 - Correlación entre las variables x e y .
 - Kurtosis.
5. Comente los resultados, resaltando los aspecto más llamativos.
6. Identifique aquellas métricas que son iguales para los conjuntos. Luego, visualice los conjuntos para caracterizar de mejor manera sus diferencias. Para esto, genere cuatro gráficos de dispersión (*scatter plots*, uno por conjunto), donde se evidencie la relación (x, y) de los pares que corresponden a cada uno.
7. Repita los pasos anteriores, esta vez descartando aleatoriamente el 10% de los datos de cada conjunto, e imputando sus valores utilizando la media de cada coordenada. Comente sobre las diferencias en los resultados.