

Actividad 2.1 – Análisis de observaciones influyentes

1. A partir del código de ejemplo utilizado en el notebook

Ejemplo_2_4_Observaciones_influyentes_Sin soluciones.ipynb

Url: https://colab.research.google.com/drive/11JM5daNQUCB_VSAOHmpFjHmFuZDsB3-i?usp=sharing

Responder a las siguientes preguntas:

- a) **(1 Punto)** Calcular la media y la mediana antes de realizar la modificación de incluir unos ingresos de 500.000€
- b) **(2 Puntos)** Aplicar el método de Probabilidad global, para detectar los outliers utilizado en el ejemplo 2_3_Outliers (**Url:** https://colab.research.google.com/drive/1C6uBUxui_Qq9ee-51vcVYcqigrSHSZNY?usp=sharing)
- c) **(1 Punto)** Repetir el mismo procedimiento de detectar los outliers para la mediana: ¿Qué ocurre?

2. A partir del código de ejemplo utilizado en el notebook

Ejemplo_2_5_Escalamiento_de_datos_Sin soluciones.ipynb

Url: <https://colab.research.google.com/drive/11vLMbjw5XmF7dJks0b04gfMdCnClE0Kw?usp=sharing>

Responder a las siguientes preguntas:

Considerar que la variable X toma los valores 1,2,3,4,5,6,7,8,9,10. Se pide:

- a) **(2 Puntos)** ¿Cuánto vale la media, mediana, la desviación estándar muestral, la varianza muestral y el rango de la variable X?
- b) **(1 Punto)** Utilizar la función describe() de Panda, para obtener la media, desviación estándar, etc...
- c) **(1 Punto)** ¿Por qué el resultado de calcular la desviación estándar con Numpy es diferente a la calculada por describe de Panda? ¿Qué ajuste sería necesario realizar para que los resultados fuesen similares/iguales?
- d) **(1 Punto)** Estandarizar la variable (escalamiento) mediante rangos y a continuación calcular la media y la mediana de la variable escalada.
- e) **(1 Punto)** Repetir el apartado anterior con el escalamiento Z - score

Formato de entrega

- Entregar un fichero en formato pdf con el siguiente contenido:
- Nombrar el archivo siguiendo el siguiente patrón:

SNS_ACT2_1_NombreApellidos.pdf

- En el documento pdf:
 - Incluir en el pie de cada página el nombre y apellidos del autor/a, así como el número de página y el total de páginas que contiene el documento.
 - Las imágenes capturadas han de tener la resolución necesaria para una buena visualización.
 - Indicar la url del GitHub donde se encuentran el/los diferentes cuadernos que has utilizado, con el objeto de consultarlos para descargarlo y verificar su funcionamiento.
 - **Importante:** Respetar la estructura de las preguntas, con el objeto de permitir una corrección homogénea para todas las actividades. De lo contrario se valorará con zeros puntos.
 - Al comienzo del notebook poner el nombre y apellidos del autor/a.