

Curso detección de anomalías
Cuestionario definitivo
Docentes: José Fernando Zea – Fernando López – Torrijos
Entrega hasta 25 de abril (11:59 pm hora colombiana)
Correo: jfzeac@unal.edu.co

1. Utilizando la metodología de Cox-Tukey lleve a cabo la identificación de proyectos de arreglos de vías terciarias con valores extremos en el total de proyectos.
Elabore un histograma de la variable transformada.
 - a. El valor de lambda es (introduzca el valor numérico):
 - b. Identifique una vez estén los valores transformados los *outliers* detectados con la regla de Tukey (utilice un valor de k igual a 1,5). El número de valores extremos identificados es:
 - c. Si repite el valor de b con k = 3. El número de valores extremos es:
2. Considere las variables *valor_sgr*, *valor_nación*, *valor_otros*, *total_proyecto* y *pctg_sgr*, lleve a cabo la identificación de valores outliers utilizando la medida de identificación de outliers “local_outlier factor”. Considere un 10% de valores anómalos **y no olvide estandarizar las variables**. Considere hasta 5 vecinos (k = 5).
 - a. El número de valores outliers identificados son: _____
 - b. El proyecto con mayor del valor LOF es:
 - c. Utilizando las mismas variables identifique los proyectos de regalías con un valor de LOF mayor a 3. El número de valores outliers identificados son: _____
3. Considere las variables *valor_sgr*, *valor_nación*, *valor_otros*, *total_proyecto* y *pctg_sgr*, lleve a cabo la identificación de valores outliers utilizando *Isolation Forest*
 - a. Por defecto, selecciona el 10% de los registros como outliers. La pregunta es: El número de proyectos que coinciden con el 10% de los proyectos con mayor LOF es: _____