

Materia:

**DISEÑO ELECTRÓNICO BASADO EN
SISTEMAS EMBEBIDOS**

Alumno:

Posadas Pérez Isaac Sayeg

Paniagua Rico Juan Julian

García Azzúa Jorge Roberto

Grado y grupo:

8°G

Profesor:

Garcia Ruiz Alejandro Humberto

Tarea 13:

Identificación de Outliers por IQR

Identificación de Outliers por IQR

1. Introducción

En el análisis de datos, la detección de valores atípicos (outliers) es un paso fundamental para mejorar la calidad de la información y evitar distorsiones en los resultados. Un outlier es un dato que se encuentra significativamente alejado del resto de los valores y puede deberse a errores de medición, ruido en los datos o eventos inusuales. Uno de los métodos más utilizados para la detección de outliers es el uso del **rango intercuartil (IQR, por sus siglas en inglés)**.

El **IQR** es una medida de dispersión estadística que permite identificar valores que se encuentran fuera de un rango esperado en una distribución de datos. A través de este método, se puede establecer un límite superior e inferior y marcar como outliers los valores que se encuentran fuera de estos límites.

2. Cálculo del IQR

El rango intercuartil (IQR) se basa en los **cuartiles**, que dividen un conjunto de datos en cuatro partes iguales. Para calcular el IQR, se siguen los siguientes pasos detallados:

1. **Ordenar los datos:** Se organiza el conjunto de datos en orden ascendente, de menor a mayor.
2. **Determinar los cuartiles:**
 - **Q1 (Primer cuartil):** Es el valor que divide el primer 25% de los datos. Se encuentra en la posición $\frac{n+1}{4}$, donde n es la cantidad total de datos.
 - **Q2 (Mediana):** Es el percentil 50, dividiendo la muestra en dos mitades iguales. Se calcula como el promedio de los dos valores centrales si el número total de datos es par, o el valor central directamente si es impar.
 - **Q3 (Tercer cuartil):** Es el valor que divide el 75% de los datos. Se encuentra en la posición $3 \times \frac{n+1}{4}$.

3. **Calcular el IQR:** Se obtiene restando Q1 a Q3:

$$IQR = Q3 - Q1$$

4. **Determinar los límites para la detección de outliers:** Los valores que caen fuera de estos límites se consideran atípicos:

$$\text{Límite inferior} = Q1 - 1.5 \times IQR \quad \text{Límite superior} = Q3 + 1.5 \times IQR$$

5. **Clasificación de los datos:** Cualquier valor menor al límite inferior o mayor al límite superior se considera un outlier y puede ser analizado para determinar su relevancia o si debe ser excluido del análisis.

Por ejemplo, si tenemos el conjunto de datos:

10,12,14,15,16,18,19,22,24,30,35,100

Siguiendo los pasos anteriores:

- **Q1** se encuentra en la posición $\frac{12+1}{4} = 3.25$, lo que nos da un Q1 aproximado de **14.5**.
- **Q3** se encuentra en la posición $3 \times \frac{12+1}{4} = 9.75$, resultando en un Q3 aproximado de **27**.
- **IQR** = $27 - 14.5 = 12.5$
- **Límite inferior** = $14.5 - (1.5 \times 12.5) = -4.25$
- **Límite superior** = $27 + (1.5 \times 12.5) = 45.75$

Por lo tanto, cualquier valor por debajo de -4.25 o por encima de 45.75 se considera un outlier. En este caso, **100** es un outlier.

3. Aplicación Práctica en Python

Para ilustrar la detección de outliers utilizando el IQR, podemos implementar un código en **Python** utilizando bibliotecas como **NumPy** y **Matplotlib**.

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Datos de ejemplo
datos = [12, 14, 15, 16, 17, 19, 21, 22, 24, 30, 35, 100] # 100 es un posible outlier

# Cálculo de Q1, Q3 e IQR
Q1 = np.percentile(datos, 25)
Q3 = np.percentile(datos, 75)
IQR = Q3 - Q1

# Cálculo de límites
limite_inferior = Q1 - 1.5 * IQR
limite_superior = Q3 + 1.5 * IQR

# Identificación de outliers
outliers = [x for x in datos if x < limite_inferior or x > limite_superior]
print("Outliers detectados:", outliers)

# Visualización con BoxPlot
plt.figure(figsize=(6,4))
sns.boxplot(x=datos)
plt.title("Identificación de Outliers por IQR")
plt.show()
```

4. Importancia en la Ciencia de Datos y Sistemas Embebidos

La detección de outliers, o valores atípicos, es un proceso esencial en diversas áreas de estudio y aplicación, especialmente en la ciencia de datos y en sistemas embebidos. En el ámbito de la ciencia de datos, la identificación de outliers contribuye significativamente a la mejora de la calidad del análisis. Los outliers pueden distorsionar los resultados de los modelos predictivos y, si no se manejan adecuadamente, pueden llevar a interpretaciones erróneas de los datos. Al aplicar técnicas de detección de outliers, los científicos de datos pueden limpiar sus conjuntos de datos, lo que resulta en análisis más precisos y decisiones mejor fundamentadas.

En el contexto de sistemas embebidos, la detección de outliers es igualmente importante, especialmente en la gestión de datos provenientes de sensores. Por ejemplo, en dispositivos como Arduino y Raspberry Pi, los sensores pueden estar

sujetos a diversas interferencias que afectan sus mediciones. La identificación de valores atípicos en estos datos permite detectar errores en las lecturas, lo que ayuda a optimizar el procesamiento de datos. Esto no solo mejora la fiabilidad de las aplicaciones, sino que también permite que los sistemas embebidos respondan de manera más efectiva a situaciones inesperadas.

En el sector financiero, la detección de outliers juega un papel crucial en la identificación de transacciones sospechosas y en la prevención de fraudes. Los sistemas de monitoreo pueden utilizar algoritmos de detección de outliers para señalar actividades inusuales que podrían indicar un comportamiento fraudulento. De este modo, las instituciones financieras pueden reaccionar rápidamente ante posibles amenazas y proteger los activos de sus clientes.

En el ámbito de la medicina, la detección de outliers es vital para identificar anomalías en señales biomédicas, como electrocardiogramas. Estas señales son fundamentales para el diagnóstico y el monitoreo de diversas condiciones de salud. Al detectar valores atípicos en los datos, los profesionales de la salud pueden identificar posibles problemas en el funcionamiento del corazón y actuar en consecuencia. En resumen, la detección de outliers es un componente esencial que mejora la calidad del análisis, optimiza el procesamiento de datos y permite la identificación temprana de problemas en múltiples campos.

5. Conclusión

El método de **Identificación de Outliers por IQR** es una técnica muy eficaz y simple para detectar valores atípicos en conjuntos de datos. Su aplicación en la ciencia de datos y en el análisis de información proveniente de sistemas embebidos permite mejorar la precisión y confiabilidad de los resultados. Con herramientas como **Python**, es posible automatizar este proceso y visualizar los datos de manera efectiva.

6. Bibliografía

1. McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media.
2. Seaborn Documentation. (n.d.). *Box plot visualization*. Recuperado de <https://seaborn.pydata.org/generated/seaborn.boxplot.html>
3. Arduino. (n.d.). *Sensor Data Analysis*. Recuperado de <https://www.arduino.cc/>