

Experiment to detect outliers in a given dataset

Description: Understand the procedure to identify outliers in a given dataset.

Soln: Aim: To detect and report outliers in a given dataset (using Interquartile Range method).

Algorithm:

- Sort the data
- Compute Q_1 , Q_2 & Q_3 (lower bound, upper bound)
- Compute $IQR = Q_3 - Q_1$

1. Sort the data

$$18 - \Sigma 2 = 18$$

2. Compute Q_1 , Q_2 & Q_3 (lower bound, upper bound)

3. Compute $IQR = Q_3 - Q_1$

(lower bound) $= Q_1 - 1.5 \times IQR$ and
(upper bound) $= Q_3 + 1.5 \times IQR$

5. Any point $<$ lower bound or $>$ upper bound is outlier and outliers points will be removed.

Non-outliers are non-outliers ("= IQR") being

lower bound, "or (all) forward bound" being

(lower-bound) " = (all) forward bound" being

(outliers) " = (all) forward bound" being

((non-outliers) for " = (all) forward bound" being

non-outliers

25/01/23

Program, displays details of trimming Q3

```

import numpy as np
import matplotlib.pyplot as plt
data = np.array([27, 50, 44, 6, 58, 61, 23, 86, 67, 20,
                 41, 35, 48, 75, 74, 61, 90, 54])
Q1 = np.percentile(data, 25)
Q2 = np.percentile(data, 50)
Q3 = np.percentile(data, 75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
outliers = data[(data < lower_bound) | (data > upper_bound)]
non_outliers = data[(data >= lower_bound) & (data <= upper_bound)]
print("Data : ", data)
print("Q1 = ", Q1, "Q2 = ", Q2, "(median.)", "Q3 = ", Q3)
print("IQR = ", IQR)
print("Lower bound (Lr) = ", lower_bound)
print("Upper bound (Ur) = ", upper_bound)
print("Outliers detected : ", list(outliers))
print("Non-outliers : ", list(non_outliers))

```

```
plt.figure()  
plt.hist(data, bins=8)  
plt.title("Histogram of data")  
plt.xlabel("Value")  
plt.ylabel("Frequency")  
plt.show()
```

```
plt.figure()  
plt.boxplot(data, vert=False)  
plt.title("Boxplot (IQR)")  
plt.xlabel("Value")  
plt.show()
```

Output:

$$Q1 = 26.0 \quad (\text{2nd quartile})$$

$$Q2 \text{ (median)} = 56.0$$

$$Q3 = 69.0$$

$$IQR = 43.0$$

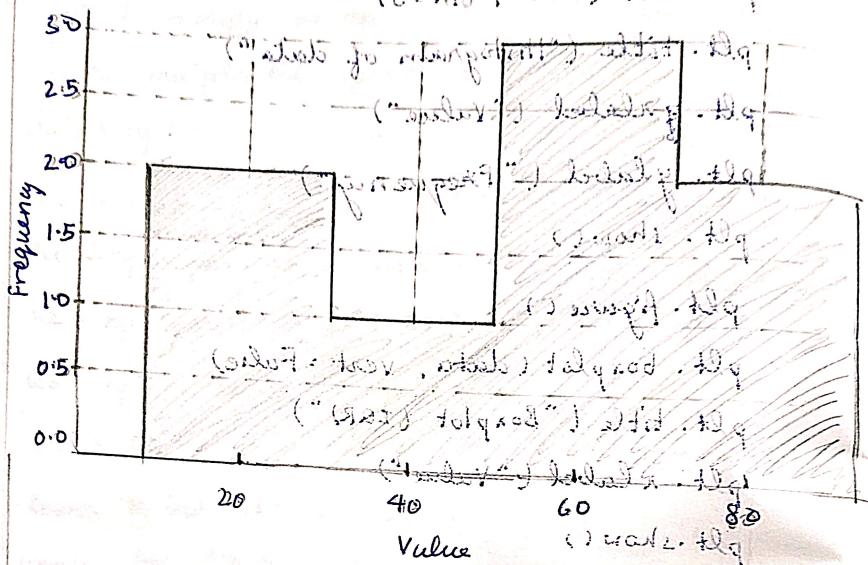
$$\text{lower bound (L)} = 38.5$$

$$\text{upper bound (U)} = 133.5$$

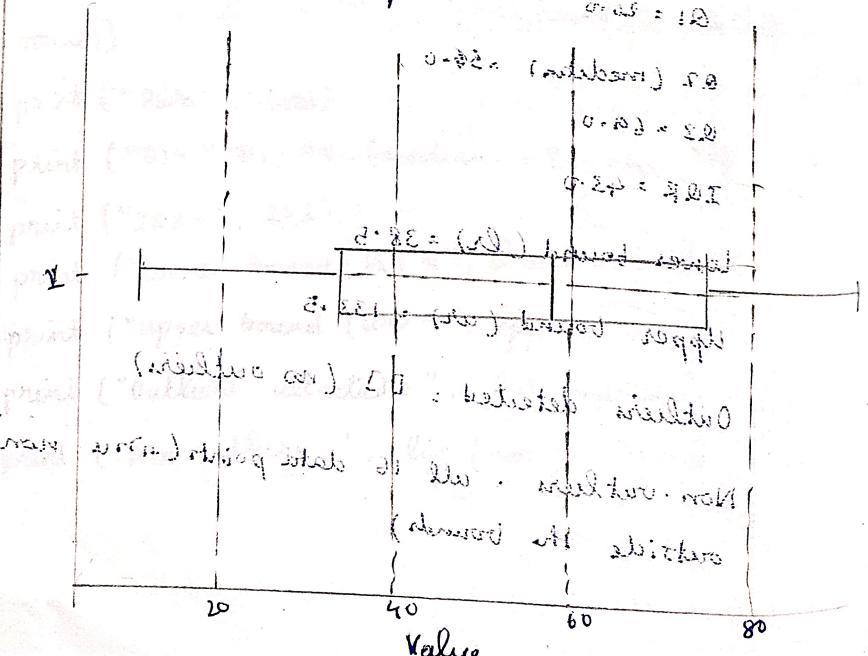
Outliers detected: [] (no outliers)

Non-outliers: all 16 data points (and none fall outside the bounds)

Histogram of Data



Boxplot (IQR)



Step 3: To find the range of outliers, we calculate the range of the data set.

In step 2, we find the range of outliers by calculating the interquartile range. We can also calculate the range of outliers by calculating the range of the data set.

Step 4: Calculate the range of outliers.

To calculate the range of outliers, we first calculate the range of the data set. Then, we subtract the range of the data set from the calculated range of outliers. This will give us the range of outliers.

We can also calculate the range of outliers by calculating the range of the data set and then subtracting the range of the data set from the calculated range of outliers.

Result:

Thus, no outliers were found, since all data points lie within the calculated bounds.