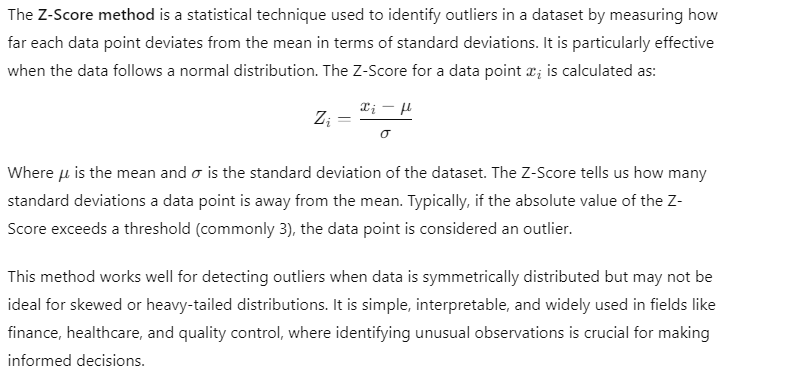
| **Name:** | **James Lewis** |
| --- | --- |
| **Roll No:** | **32** |
| **Class/Sem:** | TE/V |
| **Experiment No.:** | 6 |
| **Title:** | Implementation of outlier detection technique. |
| **Date of Performance:** |  |
| **Date of Submission:** |  |
| **Marks:** |  |
| **Sign of Faculty:** |  |

**Aim:** The aim of this experiment is to detect outliers in a dataset using the Z-Score Method, a statistical technique that identifies points that deviate significantly from the mean of the dataset.

**Objective:** To assess the effectiveness of the Z-Score method in recognizing anomalous data points in normally distributed datasets.

**Theory:**

****

**A white paper with black text

Description automatically generated**

Advantages:

* Simple to implement and interpret.
* Works well when the data is normally distributed.

Limitations:

* The Z-Score method assumes a normal distribution. For non-Gaussian distributions, this method may not be appropriate.
* Sensitive to small datasets; a few extreme values can significantly skew the mean and standard deviation.

**Code:**import numpy as np

import pandas as pd

data = {

'Values': [10, 12, 14, 15, 15, 17, 20, 50, 60, 14, 15, 16, 100]

}

df = pd.DataFrame(data)

Q1 = df['Values'].quantile(0.25)

Q3 = df['Values'].quantile(0.75)

IQR = Q3 - Q1

lower\_bound = Q1 - 1.5 \* IQR

upper\_bound = Q3 + 1.5 \* IQR

df['Outlier'] = np.where((df['Values'] < lower\_bound) | (df['Values'] > upper\_bound), True, False)

print(df)

outliers = df[df['Outlier'] == True]

print("\nOutliers detected using IQR Method:")

print(outliers)

**Output:**

Values Outlier

0 10 False

1 12 False

2 14 False

3 15 False

4 15 False

5 17 False

6 20 False

7 50 False

8 60 False

9 14 False

10 15 False

11 16 False

12 100 True

Outliers detected using IQR Method:

Values Outlier

12 100 True

**Conclusion:**   
Given a dataset of customer ages with a mean of 35 years and a standard deviation of 8 years, a customer is 60 years old. Using the Z-Score method, determine if this customer's age is an outlier with a threshold of 3. What is the Z-Score for this data point, and is it considered an outlier?

We are given:

* X=60X = 60X=60 (customer’s age)
* μ= 35μ=35 (mean)
* σ = 8σ = 8 (standard deviation)
* Threshold = 3

Z-Score formula:

Z=X−μσZ = \frac{X - \mu}{\sigma}Z=σX−μ​

Substitute values:

Z=60−35/8  
Z = 25/8  
Z=3.125

Z-Score = 3.125  
Since 3.125>33.125 > 33.125>3 (threshold), the customer’s age of 60 is considered an outlier.