**Report for IQR:**



**Summary of the Placement Dataset Analysis**

We analyzed several key statistics from the placement dataset, focusing on the Interquartile Range (IQR), the 1.5 Rule (a method for identifying outliers), and the minimum and maximum values for various attributes.

1. Interquartile Range (IQR):
   * This measures the spread of the middle 50% of the data. For example, the IQR for one of the attributes is 107.0, meaning there's a moderate variation in the middle data points.
2. 1.5 Rule:
   * This rule helps us identify outliers by extending 1.5 times the IQR beyond the first and third quartiles. For instance, using this rule, values greater than 322.0 or less than -106.0 could be considered outliers.
3. Lesser and Greater Values:
   * These represent the lower and upper bounds calculated using the 1.5 Rule. If a value is below the "Lesser" bound or above the "Greater" bound, it's likely an outlier.
4. Minimum and Maximum Values:
   * The minimum value observed in the dataset is 1, and the maximum is 940,000. These values indicate the range of data we're working with.

**Overall Summary:**

The dataset shows a wide range of values, with some potential outliers identified using the 1.5 Rule. Understanding the IQR and outliers will help us make better decisions in analyzing the placement data.

**WHU THE 1.5 RULE**

The concept of the 1.5 rule in the Interquartile Range (IQR) is commonly used in statistics to identify potential outliers in a dataset. Here's a clear explanation to help you understand why we multiply the IQR by 1.5:

### What is the Interquartile Range (IQR)?

The **Interquartile Range (IQR)** is a measure of statistical dispersion and represents the range within which the middle 50% of your data lies. It is calculated as:

IQR=Q3−Q1\{IQR} = Q3 - Q1IQR=Q3−Q1

where:

* **Q1 (First Quartile)** is the value below which 25% of the data falls.
* **Q3 (Third Quartile)** is the value below which 75% of the data falls.

### Why Use the 1.5 Rule with the IQR?

The concept of the 1.5 rule in the Interquartile Range (IQR) is commonly used in statistics to identify potential outliers in a dataset. Here's a clear explanation to help you understand why we multiply the IQR by 1.5:

### What is the Interquartile Range (IQR)?

The **Interquartile Range (IQR)** is a measure of statistical dispersion and represents the range within which the middle 50% of your data lies. It is calculated as:

IQR=Q3−Q1\{IQR} = Q3 - Q1IQR=Q3−Q1

where:

* **Q1 (First Quartile)** is the value below which 25% of the data falls.
* **Q3 (Third Quartile)** is the value below which 75% of the data falls.

### Why Use the 1.5 Rule with the IQR?

When analyzing data, it's important to identify and potentially exclude outliers because they can distort statistical analyses and affect results. The **1.5 rule** is a commonly used heuristic to detect these outliers.

How the 1.5 rule works?

1. **Calculating Boundaries**:
   * **Lower Boundary (Lesser)**: This is the minimum threshold below which a data point is considered an outlier on the lower end.
   * **Upper Boundary (Greater)**: This is the maximum threshold above which a data point is considered an outlier on the upper end.
2. **Applying the 1.5 Rule**:
   * To determine if a value is an outlier, you calculate boundaries by extending 1.5 times the IQR below Q1 and above Q3.

Lower Boundary (Lesser)=Q1−(1.5×IQR)\{Lower Boundary (Lesser)} = Q1 - (1.5 \times \{IQR})Lower Boundary (Lesser)=Q1−(1.5×IQR) Upper Boundary (Greater)=Q3+(1.5×IQR)\{Upper Boundary (Greater)} = Q3 + (1.5 \times \{IQR})Upper Boundary (Greater)=Q3+(1.5×IQR)

1. **Identifying Outliers**:
   * **Below the Lower Boundary**: Any data point that is less than Q1 - 1.5 \* IQR is considered a lower outlier.
   * **Above the Upper Boundary**: Any data point that is greater than Q3 + 1.5 \* IQR is considered an upper outlier.

The multiplier **1.5** is chosen because it balances sensitivity and specificity for outlier detection:

* **Sensitivity**: By using 1.5 times the IQR, we identify data points that are relatively far from the central 50% of the data but not too sensitive to normal variability.
* **Specificity**: It avoids being too strict, so it doesn't flag too many normal data points as outliers.

This multiplier is not universal for all types of data but works well in many cases to identify outliers without excluding too much data.

### Example for Better Understanding:

Imagine a dataset of test scores for a class:

* **Q1 = 50**
* **Q3 = 80**
* **IQR = Q3 - Q1 = 80 - 50 = 30**

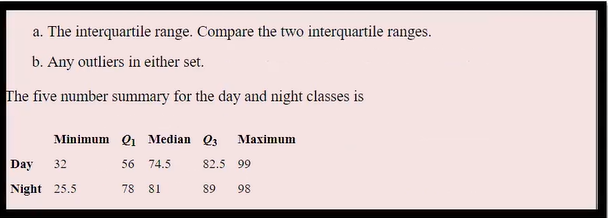
Now, using the 1.5 rule:

* **Lower Boundary**: 50−(1.5×30)=50−45=550 - (1.5 \times 30) = 50 - 45 = 550−(1.5×30)=50−45=5
* **Upper Boundary**: 80+(1.5×30)=80+45=12580 + (1.5 \times 30) = 80 + 45 = 12580+(1.5×30)=80+45=125

In this example, any test score below 5 or above 125 would be considered an outlier.

By understanding the 1.5 rule and the use of IQR, students can effectively identify outliers and understand the variability in data.

**Manual calculations**

****

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Minimum | Q1 | Median | Q3 | Maximum |
| Day | 32 | 56 | 74.5 | 82.5 | 99 |
| Night | 25.5 | 78 | 81 | 89 | 98 |

To calculate the Interquartile Range (IQR) manually for each dataset (Day and Night), follow these steps:

**Given Data**

For both Day and Night datasets:

* **Day**: Minimum = 32, Q1 = 56, Median = 74.5, Q3 = 82.5, Maximum = 99
* **Night**: Minimum = 25.5, Q1 = 78, Median = 81, Q3 = 89, Maximum = 98

**IQR Calculation**

**Calculate IQR**:

IQR=Q3−Q1

**Apply this to each dataset**:

**For Day:**

IQRDay​=Q3Day​−Q1Day​=82.5−56=26.5

**For Night:**

IQRNights​=Q3Night​−Q1Night​=89−78=11

**Summary**

* **IQR for Day**: 26.5
* **IQR for Night**: 11

**Calculate the Lesser and Greater Boundaries**

The lesser and Greater boundaries are calculated

Lesser=Q1−1.5×IQR

Greater=Q3+1.5×IQR

**Day data:**

* Lesser Boundary:

LesserDay=56−1.5×26.5=56−39.75=16.25

* Greater Boundary:

GreaterDay​=82.5+1.5×26.5=82.5+39.75=122.25

**Night Data:**

* Lesser Boundary:

LesserNight=78−1.5×11=78−16.5=61.5

* Greater Boundary:

Greater Night​=89+1.5×11=89+16.5=105.5

**Determine Outliers**

**Day Data**:

 Minimum = 32

 Maximum = 99

The boundaries of 16.25 and 122.25. **no outliers** in the Day data.

**Night Data:**

Minimum = 25.5

Maximum = 98

Final Result:

In Night Dataset has the outlier. The minimum Value 25.5 is lesser than the lesser outlier 61.50. So here we need to replace the Lesser Outlier Value 61.50 in the Night Dataset.

**Top of Form**

**Bottom of Form**

Top of Form

Bottom of Form