

## IQR – Assignment

The 1.5 IQR (Interquartile Range) rule is a commonly used method in statistics for identifying outliers in a data set. Here's how it works:

### Calculate the Interquartile Range (IQR):

Find the first quartile (Q1), which is the median of the lower half of the data set.

Find the third quartile (Q3), which is the median of the upper half of the data set.

Subtract Q1 from Q3 to get the IQR:  $IQR = Q3 - Q1$

$$IQR = Q3 - Q1$$

### Determine the Lower and Upper Bounds:

Calculate the lower bound:  $Lower\ Bound = Q1 - 1.5 \times IQR$

Calculate the upper bound:  $Upper\ Bound = Q3 + 1.5 \times IQR$

### Identify Outliers:

Any data point less than the lower bound or greater than the upper bound is considered an outlier.

### Example :

Example Data Set: 2,4,5,7,8,10,12,15,18,20,2,4,5,7,8,10,12,15,18,20

Calculate the Quartiles and IQR:

- First Quartile (Q1) = 5.5
- Third Quartile (Q3) = 15
- $IQR = Q3 - Q1 = 15 - 5.5 = 9.5$

### Lower and Upper Bounds for Different Multipliers:

#### 1. Using Multiplier of 1:

$$Lower\ Bound = Q1 - 1 * IQR = 5.5 - 9.5 = -4$$

$$Upper\ Bound = Q3 + 1 * IQR = 15 + 9.5 = 24.5$$

#### 2. Using Multiplier of 1.5:

$$Lower\ Bound = Q1 - 1.5 * IQR = 5.5 - 1.5 * 9.5 = 5.5 - 14.25 = -8.75$$

$$Upper\ Bound = Q3 + 1.5 * IQR = 15 + 1.5 * 9.5 = 15 + 14.25 = 29.25$$

#### 3. Using Multiplier of 2:

$$Lower\ Bound = Q1 - 2 * IQR = 5.5 - 2 * 9.5 = 5.5 - 19 = -13.5$$

$$Upper\ Bound = Q3 + 2 * IQR = 15 + 2 * 9.5 = 15 + 19 = 34$$

## Identify Outliers Based on Different Multipliers:

Using Multiplier of 1:

Lower Bound = -4

Upper Bound = 24.5

Data points outside the range [-4, 24.5] are considered outliers.

No data points are outside this range, so no outliers are identified.

Using Multiplier of 1.5:

Lower Bound = -8.75

Upper Bound = 29.25

Data points outside the range [-8.75, 29.25] are considered outliers.

No data points are outside this range, so no outliers are identified.

Using Multiplier of 2:

Lower Bound = -13.5

Upper Bound = 34

Data points outside the range [-13.5, 34] are considered outliers.

No data points are outside this range, so no outliers are identified.

## Summary:

- Multiplier of 1: Would identify more outliers if there were any data points just slightly beyond the interquartile range.
- Multiplier of 1.5: Is the standard choice, providing a balanced method to identify outliers.
- Multiplier of 2: Would identify fewer outliers, considering only the most extreme points as outliers.

**Table based on Given placement data:**

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
<b>Mean</b>	108.0	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
<b>Median</b>	108.0	67.0	65.0	66.0	71.0	62.0	265000.0
<b>Mode</b>	1	62.0	63.0	65.0	60.0	56.7	300000.0
<b>Q1:25%</b>	54.5	60.6	60.9	61.0	60.0	57.945	240000.0
<b>Q2:50%</b>	108.0	67.0	65.0	66.0	71.0	62.0	265000.0
<b>Q3:75%</b>	161.5	75.7	73.0	72.0	83.5	66.255	300000.0
<b>Q4:100%</b>	215.0	89.4	97.7	91.0	98.0	77.89	940000.0
<b>IQR</b>	107.0	15.1	12.1	11.0	23.5	8.31	60000.0
<b>1.5IQR</b>	160.5	22.65	18.15	16.5	35.25	12.465	90000.0
<b>LesserRange</b>	-106.0	37.95	42.75	44.5	24.75	45.48	150000.0
<b>GreaterRange</b>	322.0	98.35	91.15	88.5	118.75	78.72	390000.0
<b>Min</b>	1	40.89	37.0	50.0	50.0	51.21	200000.0
<b>Max</b>	215	89.4	97.7	91.0	98.0	77.89	940000.0

### **Analysis Summary:**

#### **IQR:**

The IQR provides a robust measure of the spread of the middle 50% of the data, which is less sensitive to extreme values compared to the range.

#### **1.5 IQR and Outlier Detection:**

Using 1.5 times the IQR is a widely accepted method for outlier detection. It establishes a threshold to identify values that are unusually far from the central part of the data.

#### **LesserRange and GreaterRange:**

These boundaries are critical for distinguishing between normal variability and potential outliers. Values falling outside these ranges are considered unusual and worthy of further investigation.

Most of the metrics, such as SSC percentage, E-test percentage, and MBA percentage, have consistent data ranges without significant outliers. HSC percentage and Degree percentage show potential outliers on both the lower and upper ends, indicating some students perform significantly differently compared to the majority.

Salary data also shows a potential outlier on the higher end, suggesting that a few individuals earn significantly more than the rest. This analysis helps identify areas where data points deviate significantly from the norm, which can be useful for further statistical analysis or decision-making processes.

### **Overall Analysis of Data Set:**

The analysis of the data set, which includes various academic percentages (SSC, HSC, Degree, E-test, MBA) and salary information, reveals consistent performance in SSC, E-test, and MBA percentages, with no significant outliers. However, HSC and Degree percentages show variability, with outliers indicating individuals with exceptionally low or high performance. The salary data also displays significant variability, with a notable high outlier.