Inter Quartile Range

Example: To find potential outlier.

a. The interquartile range. Compare the two interquartile ranges.

b. Any outliers in either set.

The five number summary for the day and night classes is

	Minimum	Q_1	Median	Q_3	Maximum
Day	32	56	74.5	82.5	99
Night	25.5	78	81	89	98

Day:

$$Q_1 = 56$$

$$Q_3 = 82.5$$

$$IQR = Q_3 - Q_1 = 82.5 - 56 = 26.5$$

$$1.5 \times IQR = (1.5) \times (26.5) = 39.75$$

Lesser Outlier:

$$Q_1 - 1.5 \times IQR = 56 - 39.75 = 16.25 //$$

Greater Outlier:

$$Q_3 + 1.5 \times IQR = 82.5 + 39.75 = 122.25 //$$

∴ No day classes is less than 16.25 and greater than 122.25.

Therefore, there is no potential outlier.

Night:

$$Q_1 = 78$$

$$Q_3 = 89$$

$$IQR = Q_3 - Q_1 = 89 - 78 = 11$$

$$1.5 * IQR = (1.5) \times (11) = 16.5$$

Lesser Outlier:

$$Q_1 - 1.5 \times IQR = 78 - 16.5 = 61.5 //$$

Greater Outlier:

$$Q_3 + 1.5 * IQR = 89 + 16.5 = 105.5 //$$

No night classes is greater than 105.5. However 61.5 is lesser outlier value and 25.5 is less than 61.5.

∴ 25.5 is a potential outlier.

Why 1.5 in formula?

The factor of 1.5 in the formula for determining outliers (specifically in the interquartile range method) is used to create a reasonable boundary for what's considered an outlier. Let me explain the reasoning behind this:

1. Basic formula: The range for non-outliers is typically defined as: [Q1 - 1.5 * IQR, Q3 + 1.5 * IQR] Where Q1 is the first quartile, Q3 is the third quartile, and IQR is the interquartile range (Q3 - Q1).
2. Choice of multiplier: The factor 1.5 was proposed by John Tukey, a famous statistician, as a compromise between being too strict and too lenient in identifying outliers.
3. Statistical properties: With normally distributed data, this 1.5 multiplier creates a range that covers approximately 99.3% of the data. This means only about 0.7% of data points would be considered outliers in a normal distribution.

4. Flexibility: The 1.5 factor provides a good balance for many datasets. It's not so strict that it flags too many points as outliers, nor so lenient that it misses obvious anomalies.
5. Customization: While 1.5 is standard, some analysts might use different multipliers (like 3) for more or less stringent outlier detection, depending on their specific needs.