

HOPE AI

IQR(Interquartile Range) Document

1) IQR: (Interquartile Range):

- To find the outliers present in the data which differs significantly from other observations.
- Measures the spread of the middle half of the data.
- It is the range for the middle 50% of the sample.
 - Formula: $IQR = Q3(75\%) - Q1(25\%)$

Outliers arise due to,

- Changes in system behavior
- Fraudulent behavior
- Human error
- Instrument error or
- Simply through natural deviations in populations

2) Why “1.5” in IQR method of outlier detection?

- It controls the sensitivity of the range and hence the decision rule.
 - A bigger scale would make the outliers to be considered as data points, while a smaller one would make some of the data points to be perceived as outliers
 - Lesser bound outliers: $Q1 - 1.5 * IQR$
 - Greater bound outliers: $Q3 + 1.5 * IQR$
-

IQR Problem to be solved:

- A. The interquartile range. Compare the two interquartile ranges.
- B. Any outliers in either set?

	Min	Q1	Median	Q3	Max
Day	32	56	74.5	82.5	99
Night	25.5	78	81	89	98

- A) The interquartile range. Compare the two interquartile ranges.

- For day:
 - The interquartile range(IQR): $Q3 - Q1 = 82.5 - 56 = 26.5$
 - For day, the IQR = 26.5
- For night:
 - The interquartile range(IQR): $Q3 - Q1 = 89 - 78 = 11$
 - For the night, the IQR = 11

- B) In order to find the outliers in the given data, initially we need to find the lesser bound outliers & greater bound outliers:

- 1) For day:

- Lesser bound outliers: $Q1 - (0.5 * IQR) = 56 - (0.5 * 26.5) = 56 - 13.25 = 42.75$
- Greater bound outliers: $Q3 + (0.5 * IQR) = 82.5 + (0.5 * 26.5) = 82.5 + 13.25 = 95.75$

- 2) For night:

- Lesser bound outliers: $Q1 - (0.5 * IQR) = 78 - (0.5 * 11) = 78 - 5.5 = 72.5$
- Greater bound outliers: $Q3 + (0.5 * IQR) = 89 + (0.5 * 11) = 89 + 5.5 = 94.5$

Result:

The obtained informations from the given data set:

	IQR value	Lesser bound outliers range	Greater bound outliers range
Day	26.5	42.75	95.75
Night	11	72.75	94.5

Hence the outliers in the given data are as follows:

From the data:

- For Day:
 - Lesser bound outlier: 25.5 (as $32 < 42.75$ (Lesser bound outliers range))
 - Greater bound outlier: 99 (as $99 > 95.75$ (Greater bound outliers range))
 - For Night:
 - Lesser bound outlier: 32 (as $32 < 72.75$ (Lesser bound outliers range))
 - Greater bound outlier: 98 (as $98 > 94.5$ (Greater bound outliers range))
-