Inter Quartile Range (IQR)

- The Interquartile Range (IQR) is a measure of statistical dispersion, representing the **difference between the 75th and 25th percentiles (Q3 and Q1) of a dataset.**
- It's often used in conjunction with a multiplier of 1.5 to identify potential outliers
- The IQR is calculated as Q3 - Q1, where Q3 is the median of the upper half of the data and Q1 is the median of the lower half.
- Essentially, it's the range within which the middle 50% of the data falls.

Why 1.5 multiplier?

- The 1.5 IQR rule **uses 1.5 as a scale to identify outliers**. It's based on the idea that data points significantly different from the rest of the dataset are outliers.
- Specifically, any data point that is more than 1.5 times the IQR below Q1 or above Q3 is considered a potential outlier.
- This rule helps to balance sensitivity (detecting actual outliers) and stringency (avoiding false positives).
- The 1.5 multiplier provides a reasonable compromise for identifying outliers while keeping the boundaries within a reasonable range.

Exercise:

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

IQR value : IQR=Q3-Q1 **Day: IQR= 82.5 – 56 = 26.5** **Night : IQR = 89 – 78 = 11**	
Lesser : (DAY) Lesser IQR = Q1-1.5*IQR =56-(1.5)*(26.5) =16.25	**Greater : (DAY)** Greater IQR = Q3+1.5*IQR =82.5+(1.5)*(26.5) =122.25
Lesser : (Night) Lesser IQR = Q1-1.5*IQR =78-(1.5)*(11) =61.5	**Greater : (Night)** Greater IQR = Q3+1.5*IQR =89+(1.5)*(11) =105.5

DAY: Range: **16.25 to 122.25** → All values (32 to 99) fall inside → So No Outliers

NIGHT: Range: **61.5 to 105.5** → But **25.5 < 61.5** → **Outlier found** at **25.5**

We can replace 25.5 in the place of 61.5 in Night classes.

Analyzing IQR values in Placement Dataset

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
Median	67.0	65.0	66.0	71.0	62.0	265000.0
Mode	62.0	63.0	65.0	60.0	56.7	300000.0
Q1:25%	60.6	60.9	61.0	60.0	57.945	240000.0
Q2:50%	67.0	65.0	66.0	71.0	62.0	265000.0
Q3:75%	75.7	73.0	72.0	83.5	66.255	300000.0
99%	87.0	91.86	83.86	97.0	76.1142	NaN
Q4:100%	89.4	97.7	91.0	98.0	77.89	940000.0
IQR	15.1	12.1	11.0	23.5	8.31	60000.0
1.5rule	22.65	18.15	16.5	35.25	12.465	90000.0
Lesser	37.95	42.75	44.5	24.75	45.48	150000.0
Greater	98.35	91.15	88.5	118.75	78.72	390000.0
Min	40.89	37.0	50.0	50.0	51.21	200000.0
Max	89.4	97.7	91.0	98.0	77.89	940000.0

SSC :
IQR = 15.1

Min = 40.89 > Lesser = 37.95
Max = 89.4 < Greater = 98.35
Conclusion: No outliers in SSC percentage data.

HSC :
IQR = 12.1

Min = 37.0 < Lesser = 42.75
Max = 97.7 > Greater =91.15
Conclusion: Two outlier exists in HSC percentage data.

Degree Percentage:
IQR = 11.0

Min = 50.0 > Lesser = 44.5
Max = 91.0 < Greater = 88.5
Conclusion: One outlier exists in degree scores.

E-test Percentage:
IQR = 23.5

Min = 50.0 > Lesser = 24.75
Max = 98.0 < Greater = 118.75
Conclusion: No outliers found in E-test scores.

MBA Percentage:
IQR = 8.31

Min = 51.21 > Lesser = 45.48
Max = 77.89 < Greater = 78.72
Conclusion: No outliers in MBA scores.

Salary:
IQR = ₹60,000

Min = ₹200,000 > Lesser = ₹150,000
Max = ₹940,000 > Greater = ₹390,000
Conclusion: There is one outlier in salary data.