

Interquartile Range



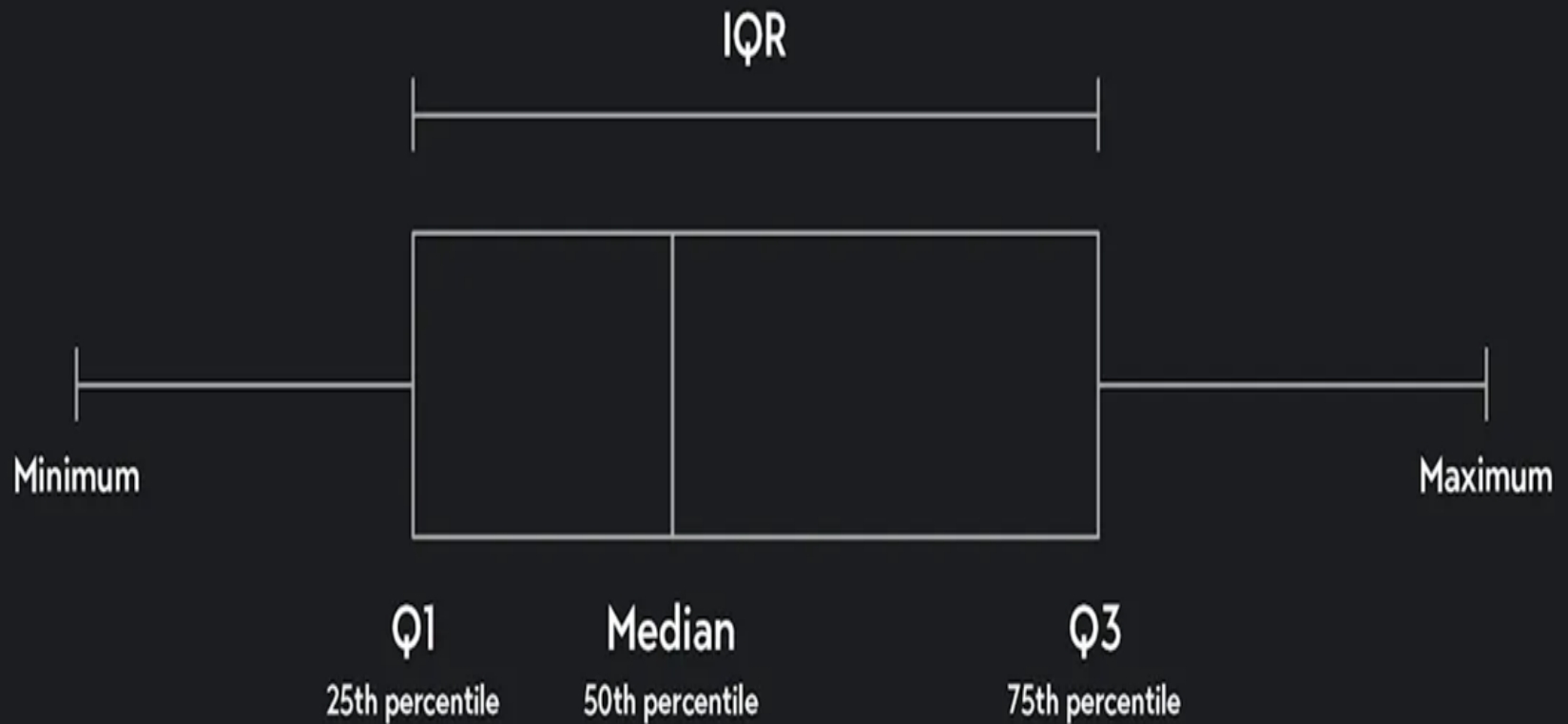
IQR

What Is the Interquartile Range (IQR)?



- ❧ In statistics, the interquartile range (IQR) is the difference between the third quartile of your data and the first quartile of your data.
- ❧ A quartile is one of three markers that divide your data into four equally sized groups, each containing roughly a quarter of your data points.
- ❧ The first quartile (also called the lower quartile or Q1) marks the 25th percentile of your data, and the third quartile (also called the upper quartile or Q3) marks the 75th percentile.

Interquartile Range = $Q3 - Q1$



Importance of Interquartile Range in Statistics



- ❧ The interquartile range is useful because it tells you how spread out the middle 50 percent of your data is. It gives you the range of values between the 25th percentile and the 75th percentile.
- ❧ The IQR is also useful as it can be used to identify outliers.

How To Calculate Interquartile Range



❧ Step 1. Count & Arrange the Data Points

- ✓ Count the number of data points and arrange them from smallest to largest. Arrange your data in ascending order from the lowest to the highest value and find the total (n) number of data points.
- ✓ If you have 10 data points, for example, $n=10$.

❧ Step 2. Calculate Q1

- ✓ To find Q1, multiply n by 25/100 (or $\frac{1}{4}$). This will give you a locator value, L.

- ✓ If L is a whole number, take the average of the L th value of the data set and the $(L+1)$ th value. This average will be your first quartile.
- ✓ If L is not a whole number, round L up to the nearest whole number and find the corresponding value in the data set.

❧ Step 3. Calculate Q3

- ✓ To find $Q3$, use the same method used to find $Q1$, except this time, multiply n by $75/100$ ($3/4$) to get the locator value, L .

❧ Step 4. Subtract Q3-Q1

- ✓ Find the interquartile range by subtracting $Q3$ from $Q1$.

Examples of How To Find the Interquartile Range



Example 1: Find the IQR for the following data set. (Notice that we have already taken the step of arranging the data in ascending order.)

10,20,30,40,50,60,70,80,90,100

Solution:

$$n=10$$

The locator value for Q_1 is $= 2.5$

$$Q_1 = 30$$

The locator value for Q_3 is $= 7.5$

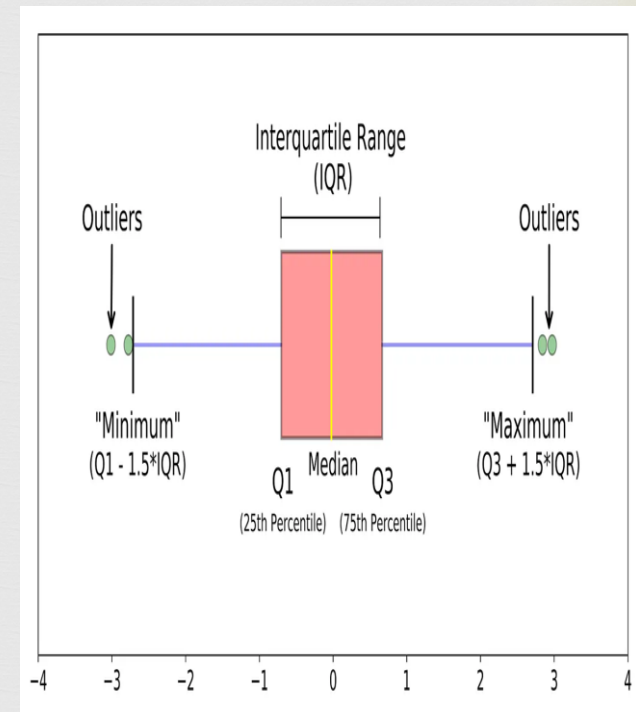
$$Q_3 = 80$$

$$\text{IQR} = 80 - 30 = 50$$

How To Find Outliers With Interquartile Range



- ✧ In addition to simply calculating the interquartile range, you can use the IQR to identify outliers in your data.
- ✧ The outlier formula also known as the 1.5 IQR rule
- ✧ Upper bound quartile - $Q3 + (1.5 \times IQR)$
- ✧ Lower bound quartile $Q1 - (1.5 \times IQR)$ as an outlier.



Why 1.5 times IQR? Why not one or two or any other number?



- ❧ The interquartile (IQR) method of outlier detection uses 1.5 as its scale to detect outliers because it most closely follows Gaussian distribution.
- ❧ As a result, the method dictates that any data point that's 1.5 points below the lower bound quartile or above the upper bound quartile is an outlier.

The Five number summary for the day and night classes is



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

Find the Day Outlier

$$\begin{aligned}\text{IQR} &= \text{Q3} - \text{Q1} \\ &= 82.5 - 56 \\ &= 26.5\end{aligned}$$

IQR is 26.5

$$\begin{aligned}\text{Lesser Outlier} &= Q1 - 1.5(\text{IQR}) \\ &= 56 - 1.5(26.5) \\ &= 56 - 39.75 \\ &= 16.25\end{aligned}$$

Lesser Outlier is 16.25. There is no lesser Outlier present in day dataset because minimum value is 32.

$$\begin{aligned}\text{Upper Outlier} &= Q3 + 1.5(\text{IQR}) \\ &= 82.5 + 1.5(26.5) \\ &= 82.5 + 39.75 \\ &= 122.25\end{aligned}$$

Upper Outlier is 122.25. There is no Upper Outlier present in day dataset because maximum value is 99.

Find the Night Outlier

$$\text{IQR} = Q3 - Q1$$

$$= 89 - 78$$

$$\text{IQR} = 11$$

$$\text{Lesser Outlier} = Q1 - 1.5(\text{IQR})$$

$$= 78 - 1.5(11)$$

$$= 78 - 16.5$$

$$= 61.5$$

Lesser Outlier is 61.5. Lesser Outlier present in night dataset because minimum value is 25.5.

$$\text{Upper Outlier} = Q3 + 1.5(\text{IQR})$$

$$= 89 + 1.5(11)$$

$$= 89 + 16.5$$

$$= 105.5$$

Upper Outlier is 105.5. There is no Upper Outlier present in night dataset because maximum value is 98.

Report for IQR of Placement dataset

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	108	67.3034	66.3332	66.3702	72.1006	62.2782	288655
Median	108	67	65	66	71	62	265000
Mode	1	62	63	65	60	56.7	300000
Q1:25%	54.5	60.6	60.9	61	60	57.945	240000
Q2:50%	108	67	65	66	71	62	265000
Q3:75%	161.5	75.7	73	72	83.5	66.255	300000
99%	212.86	87	91.86	83.86	97	76.1142	NaN
Q4:100%	215	89.4	97.7	91	98	77.89	940000
IQR	107	15.1	12.1	11	23.5	8.31	60000
1.5rule	160.5	22.65	18.15	16.5	35.25	12.465	90000
Lesser	-106	37.95	42.75	44.5	24.75	45.48	150000
Greater	322	98.35	91.15	88.5	118.75	78.72	390000
Min	1	40.89	37	50	50	51.21	200000
Max	215	89.4	97.7	91	98	77.89	940000

Below is the result of Lesser and Greater Outlier of Placement.csv file



	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Lesser	Not Present	Not Present	Present	Not Present	Not Present	Not Present	Not Present
Greater	Not Present	Not Present	Present	Present	Not Present	Not Present	Present