

	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	76.1147	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

IQR:

- The IQR represents the range of the middle 50% of a dataset.
- It's calculated as the difference between the 75th percentile (the third quartile, Q3) and the 25th percentile (the first quartile, Q1).
- Therefore: $IQR = Q3 - Q1$.

1.5 rule:

- **Calculate the IQR:** Find the difference between the third quartile (Q3) and the first quartile (Q1). ($IQR = Q3 - Q1$)
- **Determine the Boundaries:**

Lesser:

- Lower boundary: $Q1 - (1.5 * IQR)$

Greater:

- Upper boundary: $Q3 + (1.5 * IQR)$

Why 1.5 value used in iqr:

The value "1.5" is used in the IQR (Interquartile Range) method to identify outliers because it provides a good balance between sensitivity and stringency, effectively capturing significant deviations from the expected data range without being overly strict; essentially, multiplying the IQR by 1.5 creates a boundary that is roughly comparable to 3 standard deviations in a normal distribution, which is a commonly used threshold for identifying outliers.

Key points about using 1.5 with IQR:

- Practical convenience:

"1.5" is a simple, rounded number that makes calculations easier.

- Approximation to normal distribution:

While not exact, multiplying IQR by 1.5 roughly corresponds to the area considered "outlier" in a normal distribution where 3 standard deviations are used.

Middle ground:

Choosing a smaller multiplier would only identify extreme outliers, while a larger one would label too many data points as outliers.