

Create a Inter Quantile Range Analysis

IQR	107.0	15.1	12.1	11.0	23.5	8.31	60000.0
1.5rule	160.5	22.65	18.15	16.5	35.25	12.465	90000.0
Lesser	-106.0	37.95	42.75	44.5	24.75	45.48	150000.0
Greater	322.0	98.35	91.15	88.5	118.75	78.72	390000.0
Min	1.0	40.89	37.0	50.0	50.0	51.21	200000.0
Max	215.0	89.4	97.7	91.0	98.0	77.89	940000.0

IQR = $Q3 - Q1$ (spread of middle 50% of data).

1.5 Rule = $1.5 \times \text{IQR}$ (used to detect outliers).

Lesser = $Q1 - 1.5 \times \text{IQR} \rightarrow$ values below this are outliers.

Greater = $Q3 + 1.5 \times \text{IQR} \rightarrow$ values above this are outliers.

Min, Max = actual minimum and maximum values in the dataset.

1) sl_no

Lesser bound: -106 (not possible, irrelevant here).

Greater bound: 322

Actual Max: 215 \rightarrow No outliers.

2) ssc_p (Secondary %)

Lesser bound: 37.95

Greater bound: 98.35

Actual Min: 40.89 (above lesser), Max: 89.4 (below greater) \rightarrow No outliers.

3) hsc_p (Higher Secondary %)

Lesser bound: 42.75

Greater bound: 91.15

Min: 37.0 (< 42.75) \rightarrow Lower outlier present.

Max: 97.7 (> 91.15) \rightarrow Upper outlier present.

4) degree_p (Graduation %)

Lesser: 44.5

Greater: 88.5

Min: 50, Max: 91 → Upper outlier (91 > 88.5).

5) etest_p (Entrance Test %)

Lesser: 24.75

Greater: 118.75

Min: 50, Max: 98 → Within range, no outliers.

6) mba_p (MBA %)

Lesser: 45.48

Greater: 78.72

Min: 51.21, Max: 77.89 → Within range, no outliers.

7) salary

Lesser: 150,000

Greater: 390,000

Min: 200,000 (ok), Max: 940,000 (> 390,000) → Salary outliers present (very high salaries)

Why 1.5 is used in the IQR?

The **1.5×IQR rule** for outlier detection, introduced by John Tukey, is a heuristic that strikes a practical balance between being too strict (flagging too many points) and too lenient (missing true outliers).

Tukey found that **1 was too small** and **2 was too large**, so **1.5 was “just right.”** It’s not a strict law but a widely accepted standard that works well across datasets, identifying genuine unusual values without labeling normal variability as outliers.