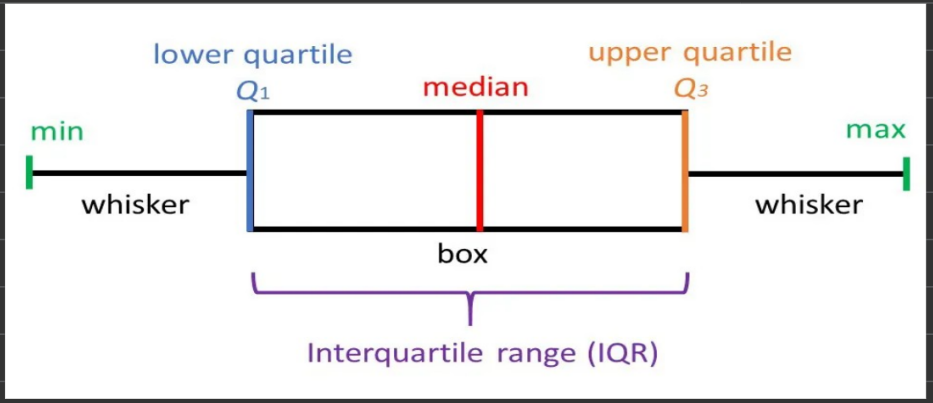

Your Ultimate Guide To Landing
Top AI roles
DECODE
AiML

5.07.4

Box Plot in Depth



→ A box plot also called whisker plot, is a powerful data visualization tool that provides summary of a numerical data distribution.

→ It uses boxes and lines to depict the data distribution.

→ Box limits indicates the central 50% data, with a central line indicating Median value.

Example : [2, 4, 6, 8, 10, 12, 14, 16]

25% 50% 25%

→ Let's understand key components of Box-Plot step by step.

① Minimum (min)
② first Quartile (Q1) → Lower Quartile
③ Median (Q2)
④ Third Quartile (Q3) → Upper Quartile
⑤ Maximum (Max)
⑥ Inter-Quartile range (IQR)
⑦ Outliers.

Median (Q2)

Q. Example : [30, 40, 50, 10, 20, 60, 70]
Find median (Q2)

↳ Sort values ⇒ [10, 20, 30, 40, 50, 60, 70]

Find the middle Element = 40

↑ median (Q2)

Hence $Q_2 = 40$

First Quartile (Q1) ← 25th Percentile

~25% of the values are below it

Q. Example : [30, 40, 50, 10, 20, 60, 70]
Find First Quartile (Q1)

↳ Sort values ⇒ [10, 20, 30, 40, 50, 60, 70]

Find the rank using formula

$$\text{Rank} = (n-1) \cdot p = (7-1) \cdot 0.25 = 1.5$$

→ So 25th Percentile lies between position 1 and 2.

↳ Interpolate between values ← **Interpolation**

$$Q_1 = x_{lower} + (\text{fraction} \times (x_{upper} - x_{lower}))$$

$$Q_1 = 20 + 0.5(30-20) = 25$$

Hence $\boxed{Q_1 = 25}$

Third Quartile (Q3) ← 75th Percentile

~75% of the values are below it

Q. Example : [30, 40, 50, 10, 20, 60, 70]
Find Third Quartile (Q3)

↳ Sort values ⇒ [10, 20, 30, 40, 50, 60, 70]

Find the rank using formula

$$\text{Rank} = (n-1) \cdot p = (7-1) \times 0.75 = 4.5$$

→ So 75th Percentile lies between position 4 and 5.

↳ Interpolate between values ← **Interpolation**

$$Q_3 = x_{lower} + (\text{fraction} \times (x_{upper} - x_{lower}))$$

$$Q_3 = 50 + (0.5 \times (60-50)) = 55$$

Hence $\boxed{Q_3 = 55}$

Inter-Quartile Range (IQR)

→ IQR describes the middle 50% of the distribution.

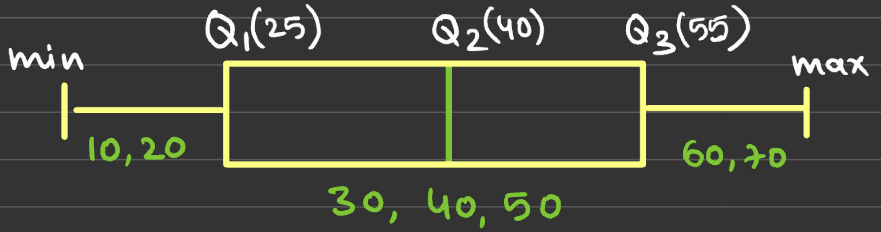
$$IQR = Q_3 - Q_1$$

→ In above Example

Example : [30, 40, 50, 10, 20, 60, 70]

On sorting values → [10, 20, 30, 40, 50, 60, 70]

Q1 → 25
Q2 → 40
Q3 → 55



NOTE :- Number of values in IQR = 3 ≠ 50%

↳ 43%

→ In large datasets, the interpolation effect is small, so IQR usually contains close to 50% of actual points.

DECODE AiML

Minimum (Min) ← Lower boundary.

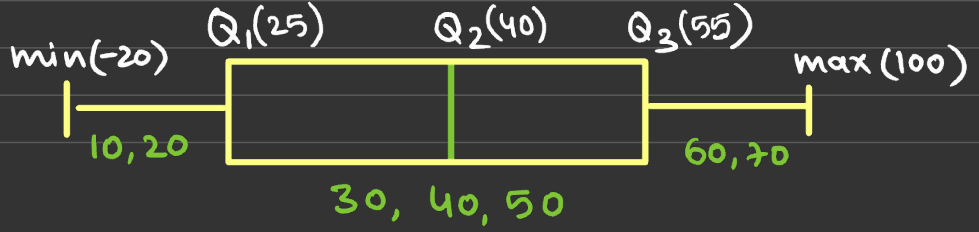
→ Minimum doesnot represents the actual minimum value of data.

$$min = Q_1 - 1.5 \times IQR$$
$$min = 25 - 1.5 \times 45 = -20$$

Maximum (Max) ← Upper boundary.

→ Maximum doesnot represents the actual maximum value of data.

$$max = Q_3 + 1.5 \times IQR$$
$$max = 55 + 1.5 \times 30 = 100$$



Outliers

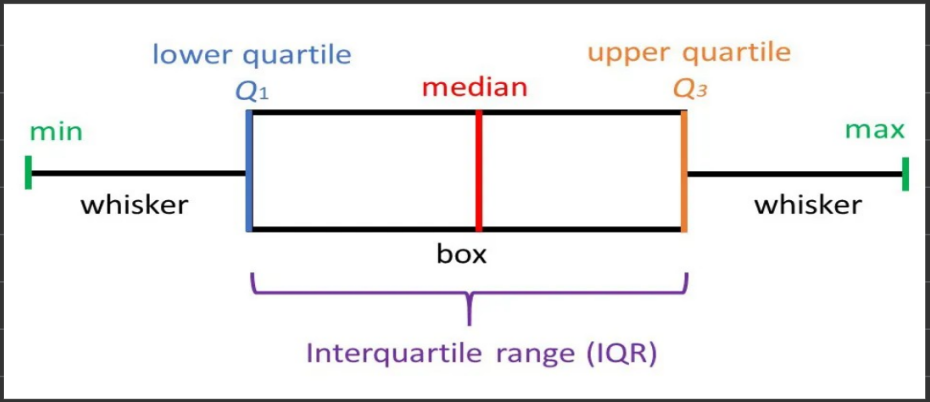
→ Any value below the lower boundary (min) and above upper boundary (max) is called Outliers.

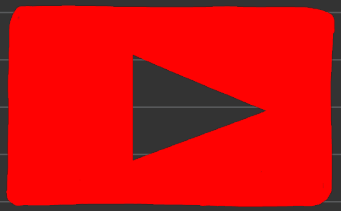
→ For above example, no value below (−20) and above max(100).

↳ So no outliers.

SUMMARY

Lecture – 5.7.2



Like  Subscribe