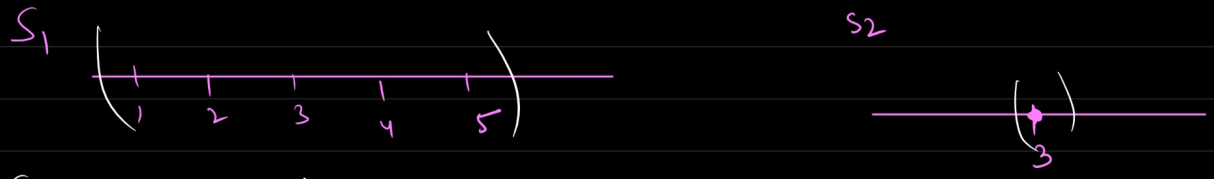
Measures of Dispersion

S_1 - 1, 2, 3, 4, 5 $\qquad$ S_2 = 3, 3, 3, 3, 3

Mean/median = 3 $\qquad$ Mean/median = 3



Can we say both S_1 & S_2 is same ?? → No

* Measures of dispersion ?

⇓

How the data is spread.

→ Range
→ Percentage & Percentile
→ Quartiles (Box plot)
→ Variance
→ Standard deviation.

* Range – difference between maximum and minimum value.

{1, 2, 3, 4, 5}

Range: 5 − 1 = 4

✓

{1, 2, 3, 4, 1000}

Range = 1000 − 1 = 999

* Outlier affects the range.

* Percentage

1, 2, 3, 4, 5

What is the percentage of nos that are odd?

$$\frac{3}{\cancel{5}} \times \overset{20}{\cancel{100}} = 60\%$$

→ 60% of the nos in the given data are odd.

* Percentile

— A percentile is a value below which a certain percentage of observation lie.

{1, 2, 3, 4, 4, 6, 7, 7, 8, 10}

* What is percentile rank of 3?

$$\text{Percentile rank of a no} = \frac{\text{No of values below that no}}{\text{Total nos (n)}} \times 100$$

$$= \frac{2}{\cancel{10}} \times \overset{10}{\cancel{100}} = 20^{th} \text{ percentile.}$$

* What value exists at 75th percentile?

$$\text{Value} = \frac{\text{Percentile} \times (n+1)}{100}$$

$$= \frac{3\cancel{75}}{\cancel{100}_{4}} \times (10+1)$$

$$= \frac{3}{4} \times 11 = 8.25^{th} \text{ no.}$$

⇓

8th number

* 75th percentile is 7.

↳ it means 75% of the no in the dat is equals to or below 7.

75th percentile ⇒ 7

8.5th ⇓ avg of 8th & 9th no.

* Quartiles

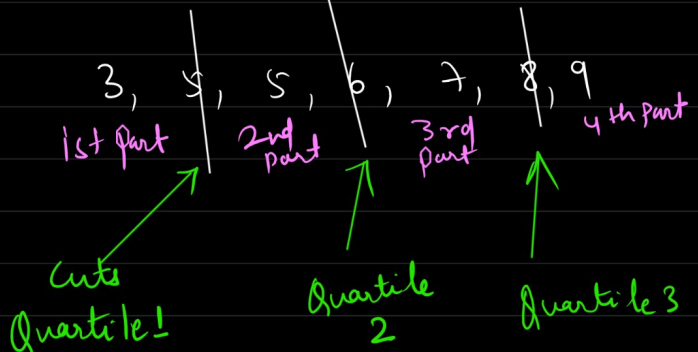
→ Quartiles are values that divides a list of numbers into quarters.

* Put the no in order
* Then cut the no into 4 equal parts
* The Quartiles are at the cut.

ex: 6, 8, 5, 5, 7, 3, 9

order → 3, 5, 5, 6, 7, 8, 9.

Cut the no into quarters



$Q_1 = 5$

$Q_2 = 6$

$Q_3 = 8$

ex: 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4

Q_1 Q_2 Q_3

total → 11 nos

if total no is odd

$$Q_1 = \frac{n+1}{4}^{th} = \frac{11+1}{4}^{th} = \frac{12}{4} = 3rd\ no$$

$Q_1 = 1$

$$Q_3 = \frac{3(n+1)}{4}^{th} = \frac{3\times(11+1)}{4}^{th} = \frac{3\times12}{4}^{th} = 9^{th}\ no$$

$Q_3 = 3$

$$Q_2 = \left(\frac{n+1}{2}\right)^{th} = \frac{11+1}{2}^{th} = \frac{12}{2}^{th} = 6^{th}\ no$$

(median)

$Q_2 = 2$

if total no is even

$$Q_1 = \frac{n}{4}^{th}\ no$$

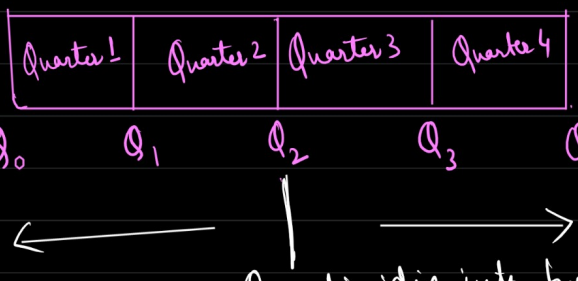
$$Q_3 = \frac{3n}{4}^{th}\ no$$

$$Q_2 = \frac{\left(\frac{n}{2}^{th}\right) + \left(\frac{n}{2}+1\right)^{th}}{2}$$

ex: 1, 2, 3, 3, 4, 4 $\quad n = 6$

$$Q_1 = \frac{6}{4}^{th} = \frac{3}{2}^{th} = 1.5^{th} = \frac{1+2}{2} = \frac{3}{2} = 1.5$$

Q_1



Q_2 dividing into two equal parts

⇓

Median

$Q_3 = \frac{3n^{th}}{4} = \frac{3\times 6^{th}}{4} = \frac{18^{th}}{4} = 4.5^{th}$

⇓ avg of 4^{th} & 5^{th} no.

$\frac{3+4}{2} = 3.5 \Rightarrow Q_3$

$Q_2 = \frac{\frac{6}{2}^{th} + \left(\frac{6}{2}+1\right)^{th}}{2}$

$= \frac{3rd + 4^{th}}{2} = \frac{3+3}{2} = 3$

Q_2

Five point Summary

1, 1, 1, 1, 2, 2, 2, 3, 3, 3

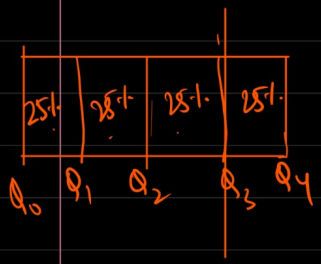
Q_0 (min) — 1 → min → 0% — 0^{th} percentile

Q_1 — 1 → 25% → 25^{th} percentile

Q_2 — 2 → 50% → 50^{th} percentile (median)

Q_3 — 3 → 75% → 75^{th} percentile

Q_4 (maximum) — 4 → 100% → 100% percentile



transaction amount

1000 → min = 1000

2000

3000 → $Q_1 = 5000$ (25^{th} percentile) → 25% of transaction amount is equals to or below 5000 in the data.

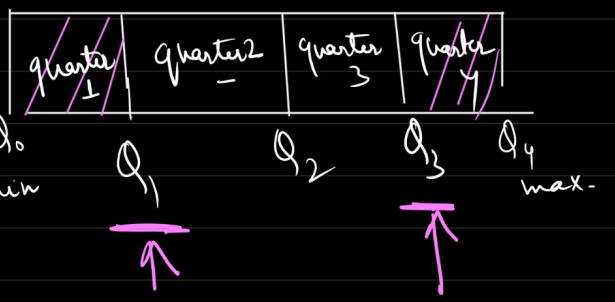
→ $Q_2 = 10000$ → 50% of transaction is equals to or below 10000

→ Q_3 → 75^{th} percentile

→ Q_4 → max.

Disadvantage of range → Outlier affects the range.

max − min
$(Q_4) - (Q_0)$
→ maximum & min value will lie in Quarter 1



* Inter quartile Range

$$IQR = Q_3 - Q_1$$

(since IQR deals with Q_3 & Q_1, it is not affected by outliers.

2, 4, 4, 5, 6, 7, 8

Q_1 ↑ , Q_3 ↑

$IQR = 7 - 4 = 3$.

outlier ↓

{2, 3, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 8, 99}

N = 16

Q_1 ↑ , Q_3 ↑

$Q_1 = 25^{th}$ percentile $= \frac{25}{100} \times 16^{th} = \frac{1}{4} \times 16^{th} = 4^{th}$ no

$Q_3 = 75^{th}$ percentile $= \frac{75}{100} \times 16^{th} = \frac{3}{4} \times 16^{th} = 12^{th}$ no

Q_2 = avg of 8^{th} & 9^{th} no $= \frac{5+5}{2} = 5$.

$Q_0 = 2$

$Q_1 = 3$

$Q_2 = 5$

$Q_3 = 6$

$Q_4 = 99$

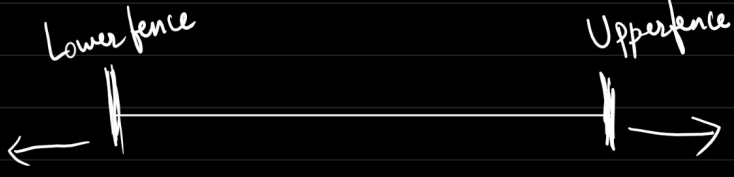
$IQR = Q_3 - Q_1 = 6 - 3 = 3$

* Use case of IQR → To detect Outliers

Outlier is extreme values

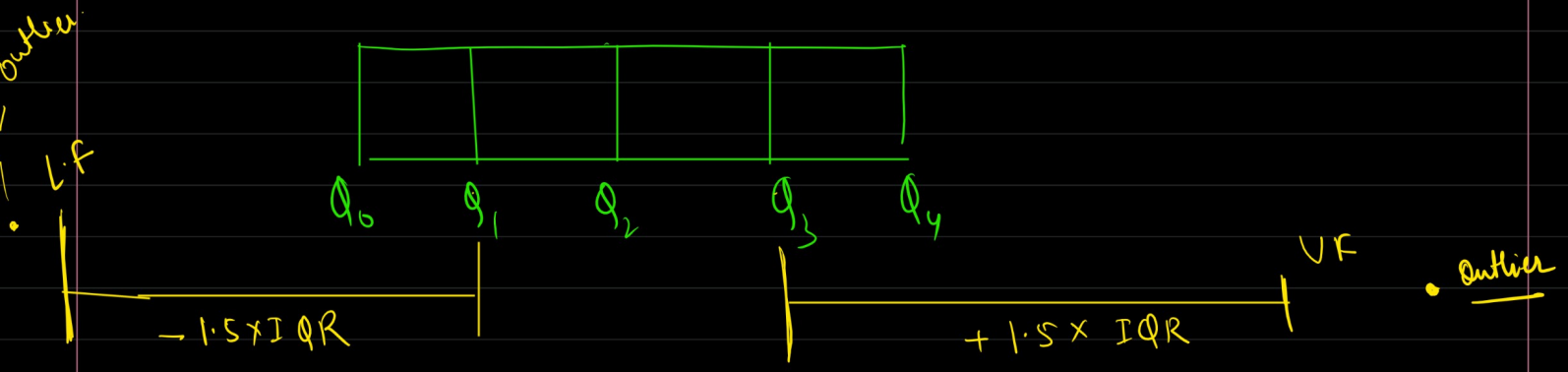
Lowerfence :- $Q_1 - 1.5 \times IQR$

Upperfence $= Q_3 + 1.5\ IQR$



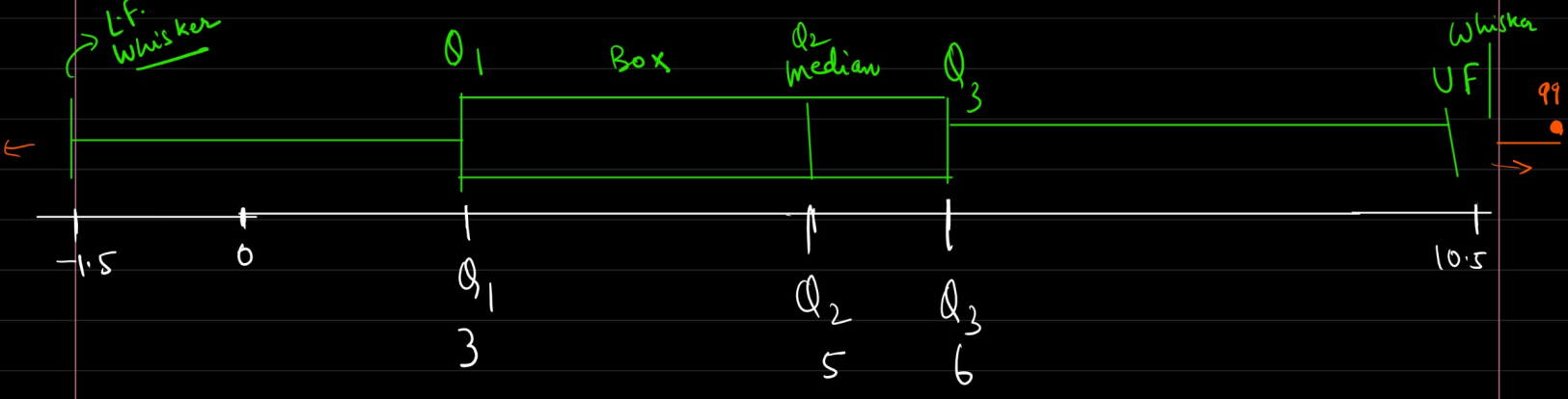
$LF = 3 - 1.5 \times 3 = -1.5$

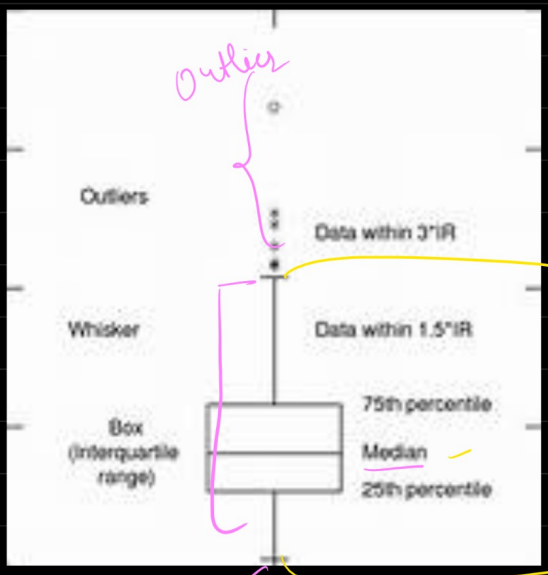
$UF = 6 + 1.5 \times 3 = 4.5 + 6 = 10.5$



* Box-Whisker plot → To detect outlier

2, 3, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 8, 99





outlier

upperfence

lowerfence

outlier