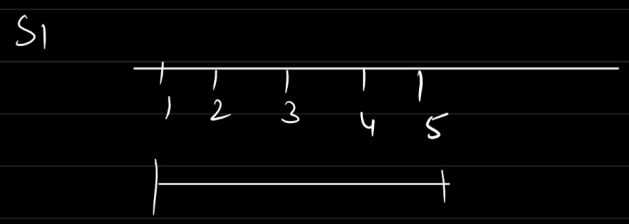
* Measures of Dispersion

S_1 – 1, 2, 3, 4, 5

mean/median – 3

S_2 = 3, 3, 3, 3, 3

mean/median = 3



* How the data is spread?

→ Range
→ Percentage and percentile
→ Quartiles (Boxplot)
→ 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 and percentiles

1, 2, 3, 4, 5

What is the percentage of nos that are odd?

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

* Percentile

Defⁿ – 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 the percentile rank of 3 ?

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

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

what value exists at 75^{th} percentile?

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

$$= \frac{75}{100} \times (10+1)$$

$$= \frac{3}{4} \times 11 = 8.25^{th} \text{ number}$$

$\Downarrow$

8^{th} number

(8.5^{th})

$\Downarrow$

avg of 8^{th} & 9^{th} no

75^{th} percentile – 7.

→ 75^{th} percentile is 8

↳ It means 75% of the no in the data is equals to or below 8.

* Quartile

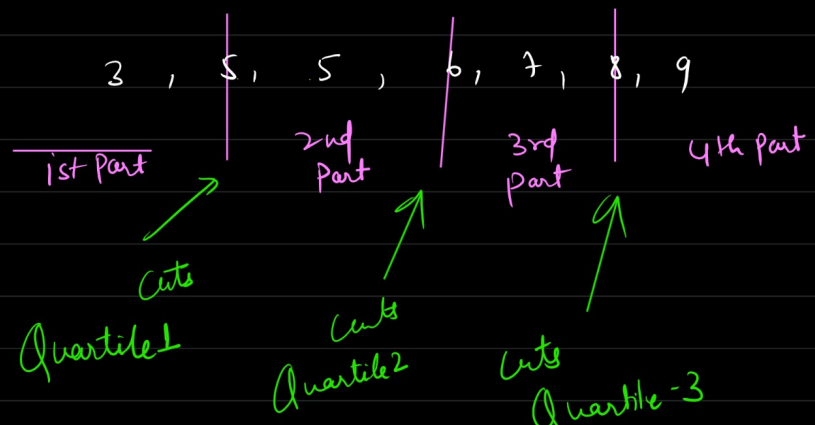
→ Quartiles are values that divides a list of numbers into quarters.

* Put the no in order
* then cut the number 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.



$Q1 \rightarrow 5$

$Q2 \rightarrow 6$

$Q3 \rightarrow 8$

Ex - 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4

total - 11 nos. Q_1 Q_2 Q_3

if total no is odd

- $Q_1 = \frac{n+1}{4}^{th} = \frac{11+1}{4}^{th} = \frac{12}{4}^{th} =$ 3rd no. $Q_1 = 3$
- $Q_3 = \frac{3(n+1)}{4}^{th} = \frac{3(11+1)}{4}^{th} = \frac{3\times12}{4} = 9^{th}$ no $Q_3 = 3$
- $Q_2 = \left(\frac{n+1}{2}\right)^{th} = \frac{11+1}{2}^{th} = 6^{th}$ no. $Q_2 = 2$ (median)

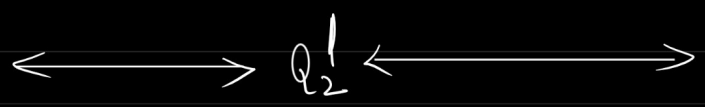
if total no is even

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

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

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

1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4



Median - the point that divides the data into two equal parts

* even no. of nos

1, 2, 3, 3, 4, 4 $\quad N = 6$

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

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

4^{th} & 5^{th} no

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

Q_2 (median when the total nos are even)

$\text{avg}\left(\frac{n}{2}^{th} \text{ no}, \frac{n}{2}^{th} + 1\right)$

$\text{avg}\left(\frac{6}{2}^{th}, \frac{6}{2}^{th} + 1\right)$

$\text{avg}\left(3^{rd}, 4^{th}\right)$

$= \frac{3+3}{2} \Rightarrow 3 \quad Q_2 = 3$

Five point summary

Q_0 (min) $- 1 \longrightarrow 0\% - 0^{th}$ Percentile

$Q_1 - 1 \longrightarrow 25\% - 25^{th}$ percentile

(median) $Q_2 \longrightarrow 2 \longrightarrow 50\% - 50^{th}$ percentile

$Q_3 - 3 \longrightarrow 75\% - 75^{th}$ percentile

Q_4 (maximum) $- 4 \longrightarrow 100\% - 100\%$ Percentile

transaction amount

1000
2000
3000
.
.
.

min = 1000 (Q_0)

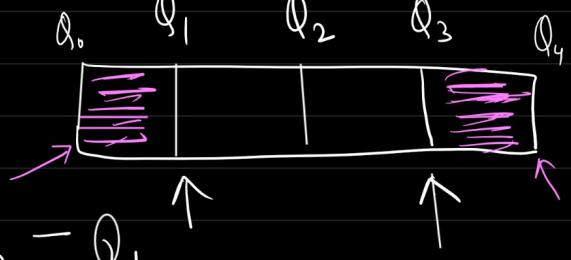
$Q_1 = 25^{th}$ percentile

↳ 5000 → 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 – 75th percentile
Q_4 – max no.

Inter Quartile range



$IQR \Rightarrow Q_3 - Q_1$

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

Q_1 = 4, Q_3 = 7

$IQR = 7 - 4 = 3$

{2, 3, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 8, 99} → Outlier.

Q_1, Q_3, $N = 16$

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

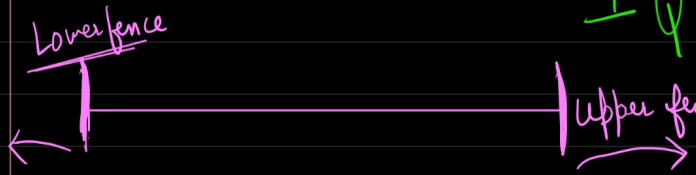
$Q_3 = 75^{th}$ percentile $= \frac{75}{100} \times 16^{th} = \frac{3}{4} \times 16^{th} = 12^{th}$ no.

Q_2 = avg of 8^{th} & 9^{th} no $= 5$

$IQR = 6 - 3 = 3$

$(Q_3) - (Q_1)$

Outliers are extremes value



$$\text{Lower fence} = Q_1 - 1.5 \times IQR$$
$$\text{Upper fence} = Q_3 + 1.5 \times IQR$$

$L.F = 3 - 1.5 \times 3 = -1.5$

$U.F = 6 + 1.5 \times 3 = 4.5 + 6 = 10.5$

Box – whisker plot

{2, 3, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 8, 9}