

Descriptive Statistics

Agenda:-

① Descriptive vs Inferential stats

① (A) Measure of Central tendency

→ Mean

→ median

→ mode

② (B) Measure of variability

→ Range

→ variance

→ std. dev

③ (C) weighted avg

④ (D) IQR

→ Real world data

⑤ (E) Random variable

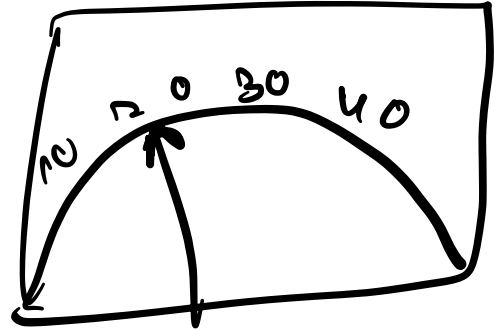
⑥ (F) Distribution function.

* Statistics

Descriptive

↳ describing

20 km/hr



Inferential

↳ Inference / Prediction / Conclusion

* Measure of Central tendency

- ① Mean
- ② Median
- ③ mode.

Eg:- 30L, 30L, 35L, 40L, 40L

$$\begin{array}{l} \text{Avg /} \\ \text{Mean} = \\ \text{ } \rightarrow \text{ } (\mu) \end{array} = \frac{\overset{x_1}{30} + \overset{x_2}{30} + \overset{x_3}{35} + \overset{x_4}{40} + \overset{x_5}{40}}{5}$$

$$= \underline{\underline{35L}}$$

$$\mu = \frac{\sum_{i=1}^N x_i}{N}, \text{ where } \underline{\underline{N=5}}$$

Debias data

→ 3 cr / yr

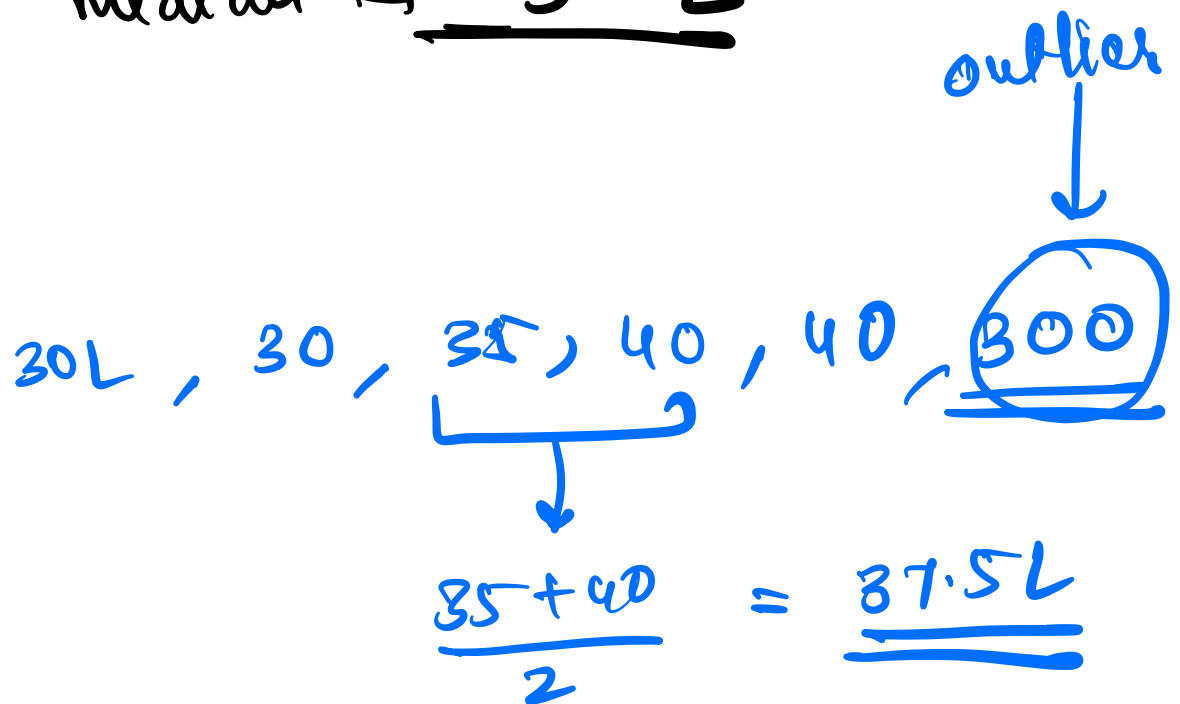
New mean =

$$= \frac{x_1 + x_2 + x_3 + x_4 + x_5}{6} + \text{outlier}$$

30 + 30 + 35 + 40 + 40 + 800L

$$\mu = \underline{\underline{79L}}$$

* Median :-



(*) Median is more Robust to outliers.

* Mode

90, 90, 90, ~~80~~, 90, 70, 95, 90

mode: 90 (5)

Q

2, 2, 3, 3, 4 } Bimodal
2 2 1

Q1 :-

4th age $\rightarrow x$

$$\frac{20 + 22 + 28 + x}{4} = 24$$

$$\underline{\underline{x = 26}}$$

20 2 2 2 6 28
 └──────────┘

Median :-

$$\frac{22 + 26}{2} = \underline{\underline{24 \text{ yrs}}}$$

* Weighted Avg :-

CGPA
or
GPA

Subj	Credit	Grade
Maths	<u>3</u>	<u>5</u>
History	4	4
Chem	3	5
Eng	2	3
=		<u>12</u>

Maths $\rightarrow 3 \times 5 = 15$

Hist $\rightarrow 4 \times 4 = 16$

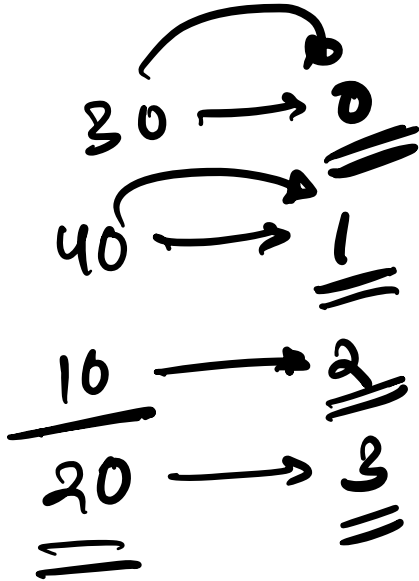
Chem $\rightarrow 3 \times 5 = 15$

$$\text{Eng} \rightarrow 2 \times 3 = \frac{6}{52}$$

$$\text{GPA} = \frac{52}{12} = \underline{\underline{4.33}}$$

(8)3

Total \rightarrow 100



0 + 0 + 0 + 0 , 1 , 1 , 1 , 2 , 3 , ...

100

$$= 30 \times 0 + 40 \times 1 + 10 \times 2 + 20 \times 3$$

100

$$= \boxed{1.2}$$

Q.4)

mean $2C \longrightarrow 40 \text{ kg.}$

mean $(2C + M) \longrightarrow \underline{45 \text{ kg}}$

$M = ?$

$$\text{mean} = \frac{\text{Sum}(2C)}{2}$$

$$40 = \frac{\text{Sum}(2C)}{2}$$

$$\text{Sum}(2C) = \underline{\underline{40 \times 2}}$$

$$\frac{\overset{c_1 + c_2}{\uparrow} \text{sum}(2c) + M}{3} = 45$$

$$\text{sum}(2c) + M = 45 \times 3$$

$$40 \times 2 + M = 45 \times 3$$

$$M = 45 \times 3 - 40 \times 2$$

$$= 135 - 80$$

$$M = 55$$

* Measure of variability

① Range

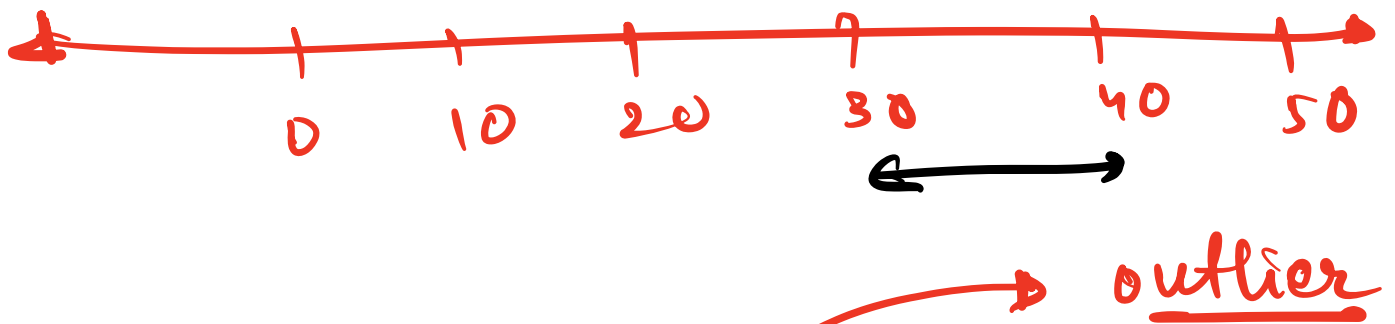
↳ max ← min

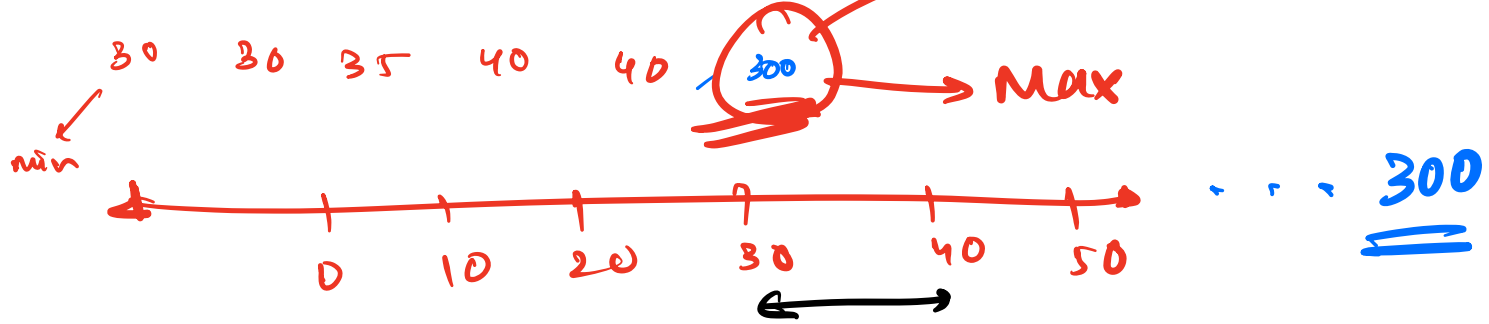
min 30 30 35 40 40 max

Range = Max - min

$$= 40 - 30$$

10 ✓





$$\text{New range} = \text{max} - \text{min}$$

$$= 300 - 30$$

$$= \boxed{270}$$

IOE

* Random Variable (RV)

→ Scenarios :-

→ weather forecast

→ Stock mkt prediction

→ Coin Toss

① Discrete RV

② Continuous RV

① Discrete RV

1) Coin Toss $\rightarrow \{ \underline{H}, \underline{T} \}$

2) Dice Throw $\rightarrow \{ \underline{1}, \underline{2}, \underline{3}, \underline{4}, \underline{5}, \underline{6} \}$

2) Continuous RV

\rightarrow Height of a person

5 ft $\xrightarrow{\quad \quad \quad \infty \quad \quad \quad}$ 6 ft

5.01 ft

5.125 ft

.....

* Distribution functions

① Probability Density function.
(PDF)

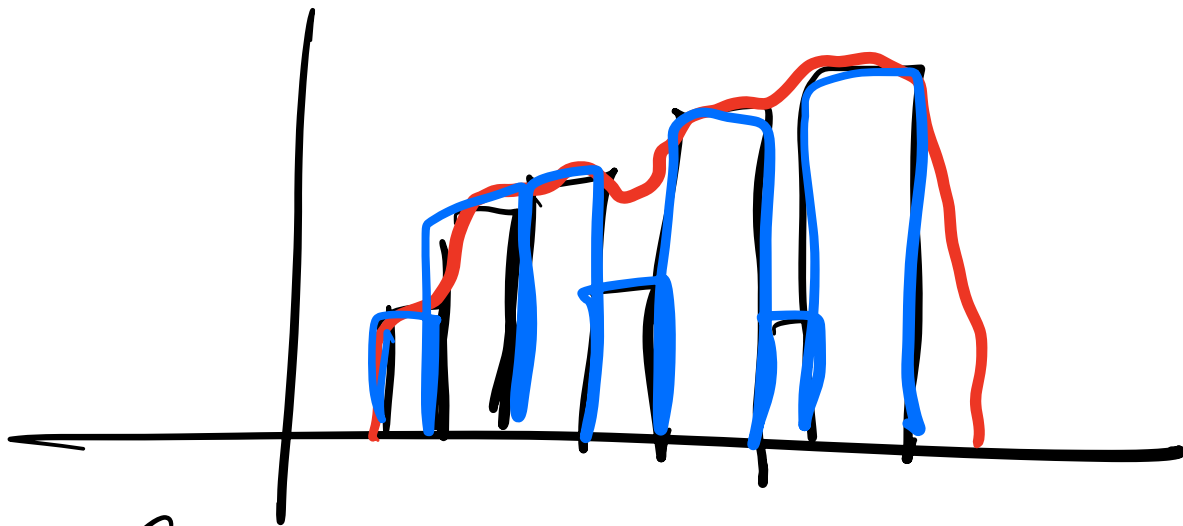
Continuous RV →

② Probability mass function
(PMF)

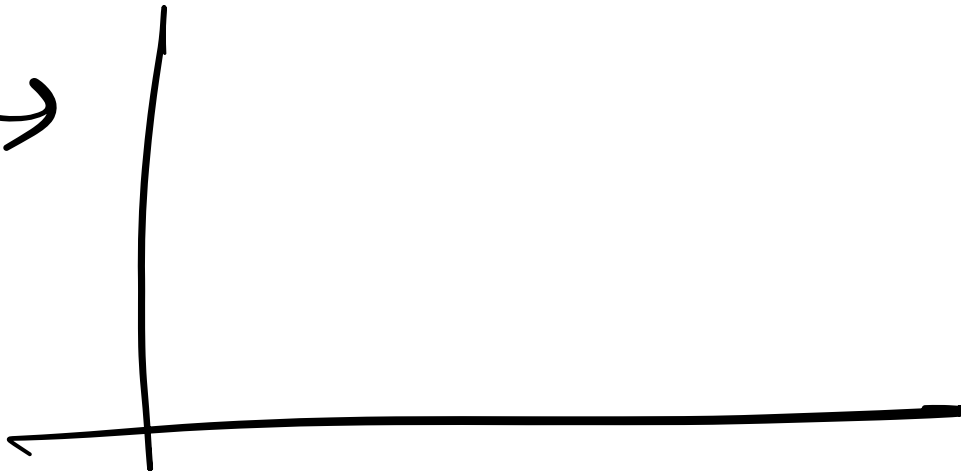
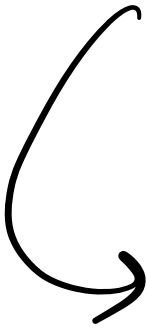
Discrete RV →

③ CDF → Cumulative Distribution function
↳ Continuous + Discrete RV

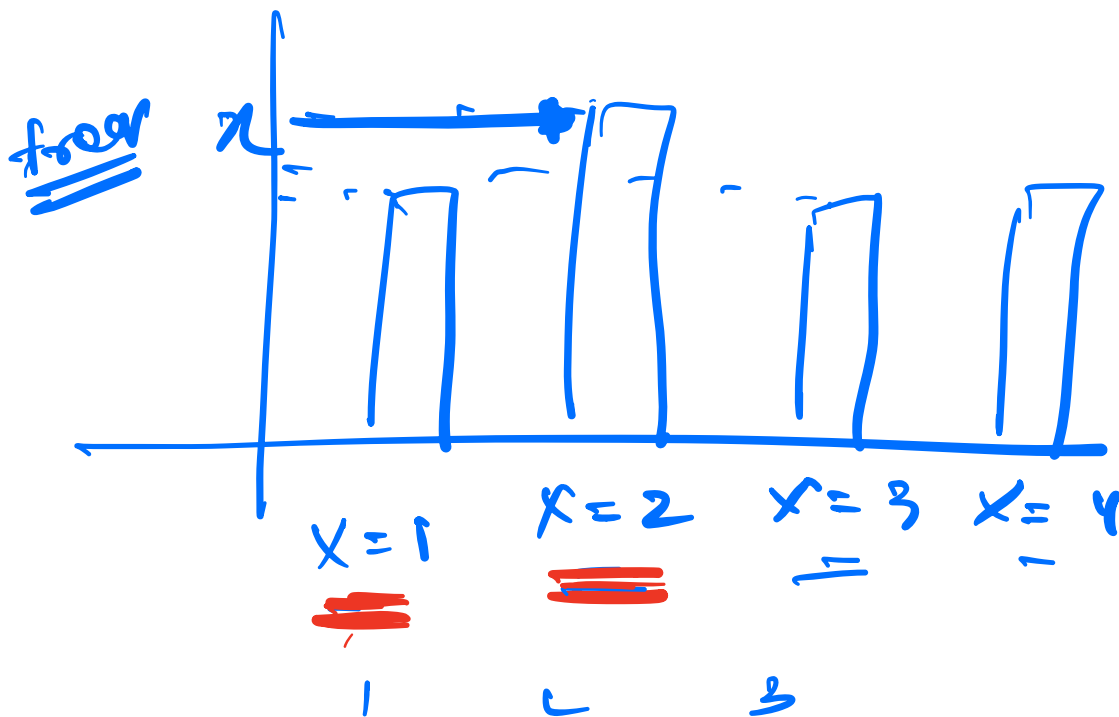
* PDF : Continuous



Continuous



* PMF : Discrete



CDF:-

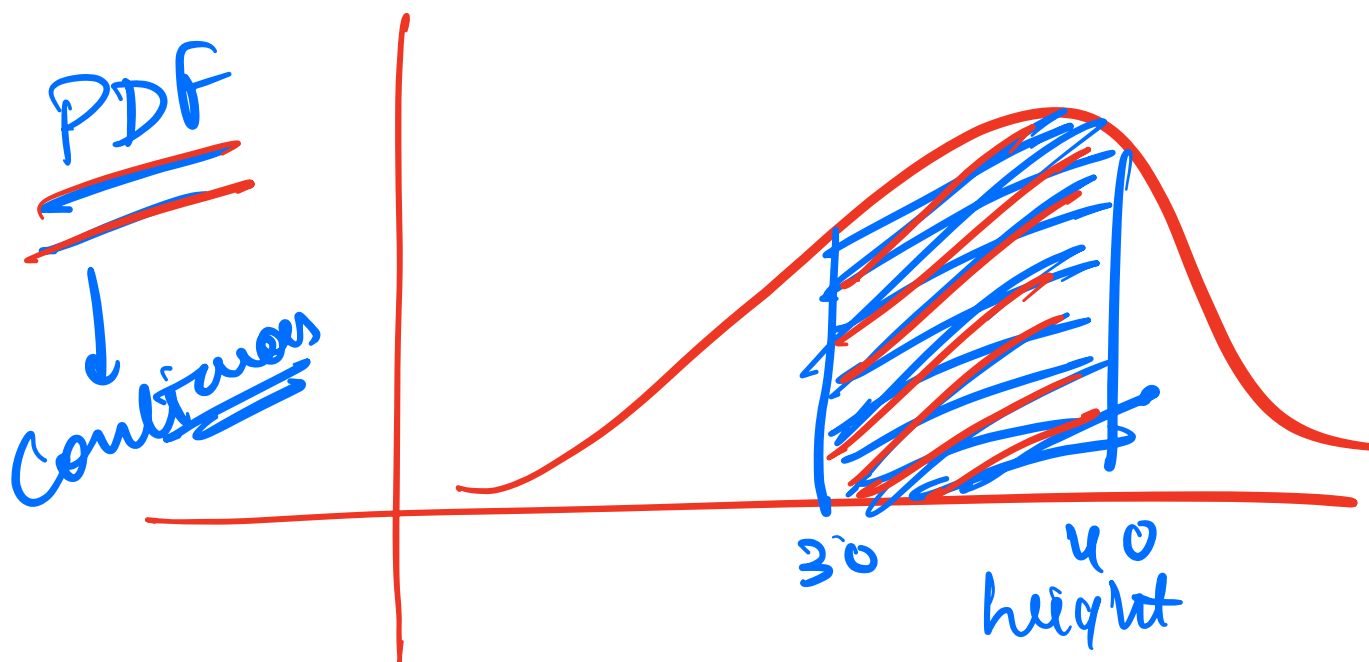
↓

$$\underline{P(X < 3)}$$

(Dice)

$$= \underline{P(X=1)} + \underline{P(X=2)}$$

PMF PMF



$$P(30 \leq x \leq 40)$$

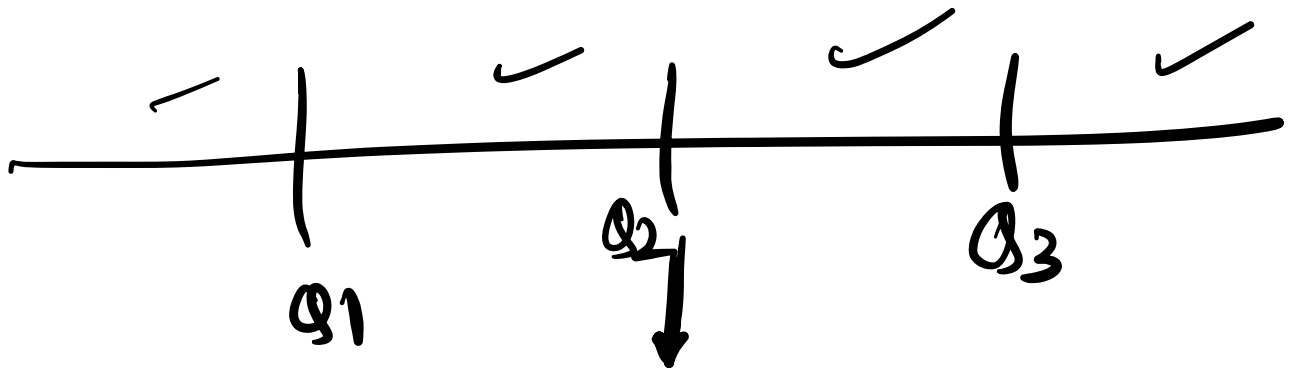


IQR:

↳ Inter Quartile Range

Quartile

↳ 4 equal parts

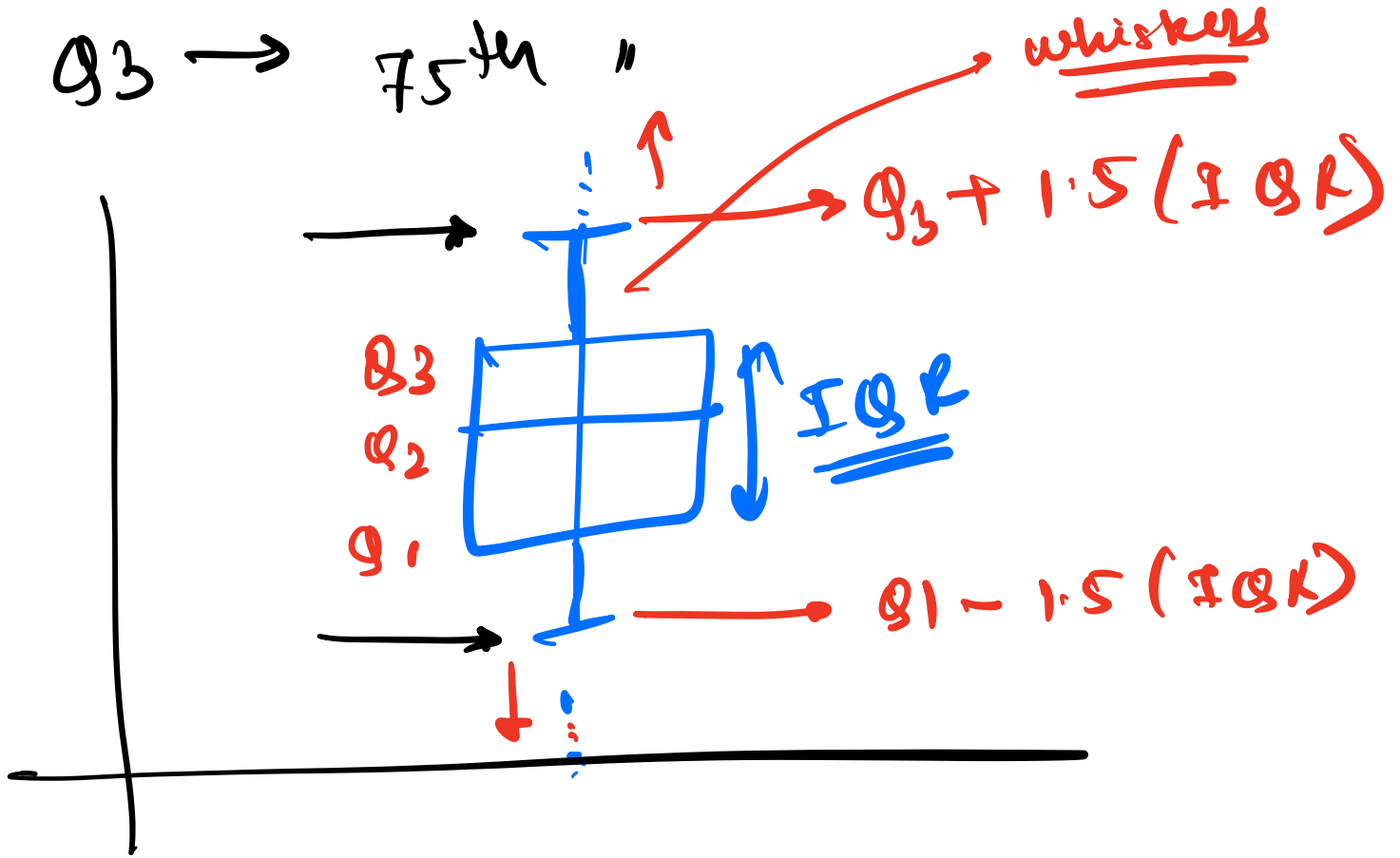


Median
(50th percentile)

$Q_1 \rightarrow 25^{\text{th}}$ percentile

$Q_2 \rightarrow 50^{\text{th}}$ (median)

93 \rightarrow 75th " whiskers



$$IQR = [\underline{q_3} - \underline{q_1}]$$