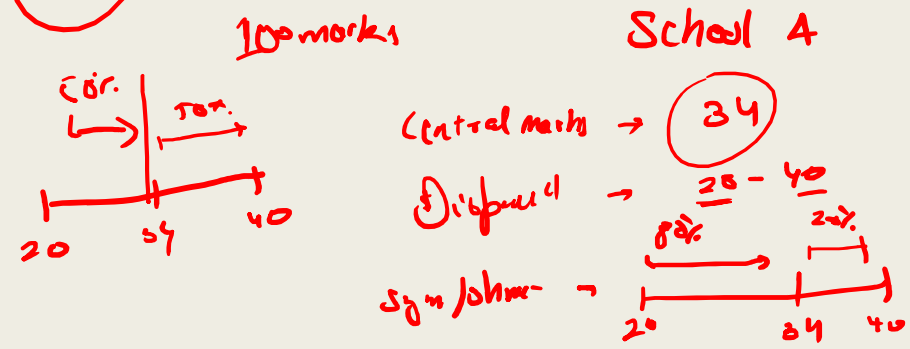
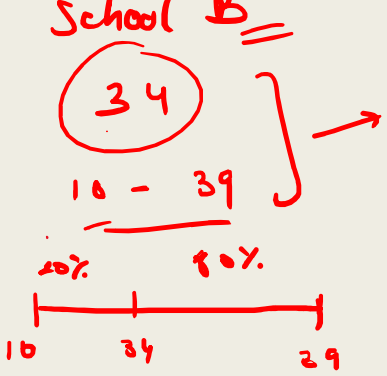
Descriptive Statistics

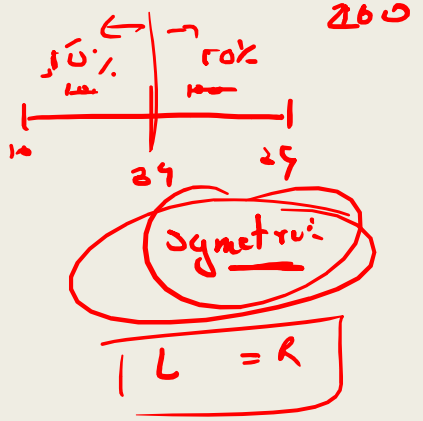
deals with the processing of data without attempting to draw any inferences from it. The characteristics of the data are described in simple terms. Events that are dealt with include everyday happenings such as accidents, prices of goods, business, incomes, epidemics, sports data, population data.

When we give description of data, there can be 3 kinds:

1. Measures of Central Tendency – Mean, Median and Mode
2. Measures of Dispersion – Standard Deviation, Variance, Range, IQR (Inter Quartile Range)
3. Measure of Symmetricity/Shape – Skewness and Kurtosis

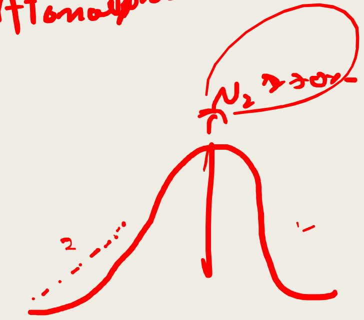




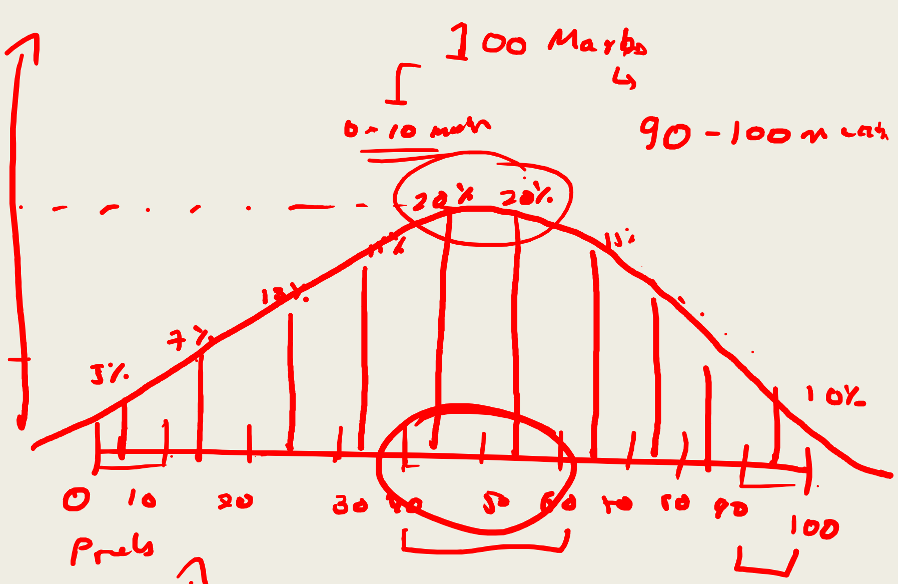


→ Central Tendency →

Atmosphere



70%



500 students
500
300
100
50
10%
25

Central Tendency

Avg

Prob

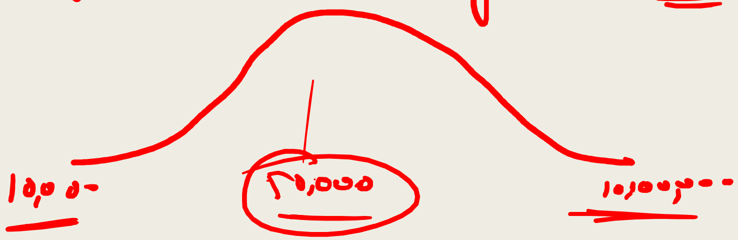
→ Normal Distribution

Natural Distributions

Bell-curve

↳ Probability dist.

Has the highest Prob to occur in your data



Employee

→ Features

1. Measure of Central Tendency

A measure of central tendency is a summary statistic that represents the centre point or typical value of a dataset. These measures indicate where most values in a distribution fall and are also referred to as the central location of a distribution.

1. Mean

Average value of the set of Numbers. Mean is a a number around which a whole data is spread out. Denoted by μ for population mean and for sample mean. → $\bar{x}$ (x-bar)

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{x}$$

Example: Find the mean of 5,5,2,6,3,8,9?
A: Mean is (5+5+2+6+3+8+9) / 7 = 38/7 = 5.43

2. Median

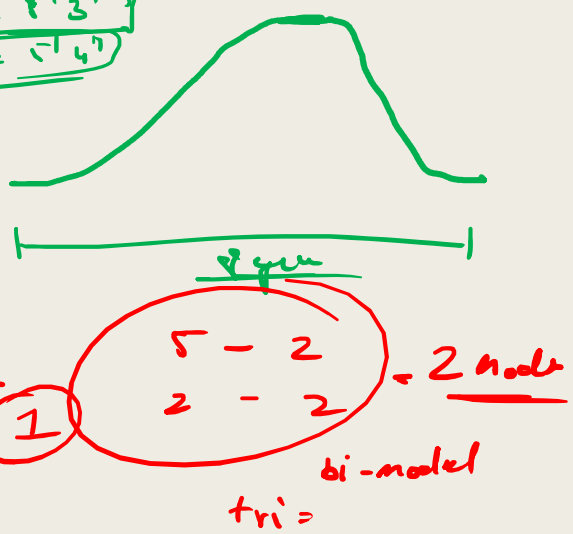
Median is the value *which divides the data in 2 equal parts* i.e. number of terms on right side of it is same as number of terms on left side of it when data is arranged in either **ascending or descending order**.
(**Note**: If you sort data in descending order, it won't affect median but IQR will be negative. IQR will be discussed in next slide.)

Example: Find the Median of 5,5,2,6,3,8,9?
A: Putting it in ascending order = 2,3,5,5,6,8,9. Hence, Median = Mid Number = 5.
(**Note**: Median of a even set of numbers can be found by taking the average of the 2 middle numbers.
E.g. Median of 2,3,4,7 = average of (3 and 4) = 3.5)

3. Mode

Mode is the term *appearing maximum time* in data set i.e. term that has highest frequency.

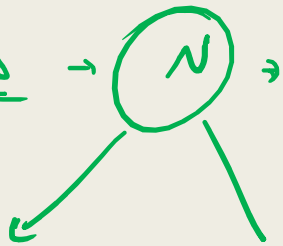
Example: Find the Median of 5,5,2,6,3,8,9?
A: Mode = Maximum number of repetition in dataset = 5. Hence, Mode = 5.
(**Note**: If there is no repetition of data then mode is not present. E.g.: What is the mode of 1,2,3,5,6?
A: None i.e. mode is not present.)



Median → center by term / position

1 → Sort your data → Asc (low to high) / Desc (high to low)

2 → calculate the # (number) of observations (count) → N → how many terms (values) do you have?!



Even (N)

→ Avg ($\frac{N}{2}^{th}$ term, $\frac{N}{2}+1^{th}$ term)

↓

median

$N = 12$, $\frac{N}{2} = 6$

$\frac{N}{2}+1 = 7$

6^{th} term

7^{th} term

$\frac{6^{th}\text{ term} + 7^{th}\text{ term}}{2}$ = median

10, 12, 1053

Odd (N)

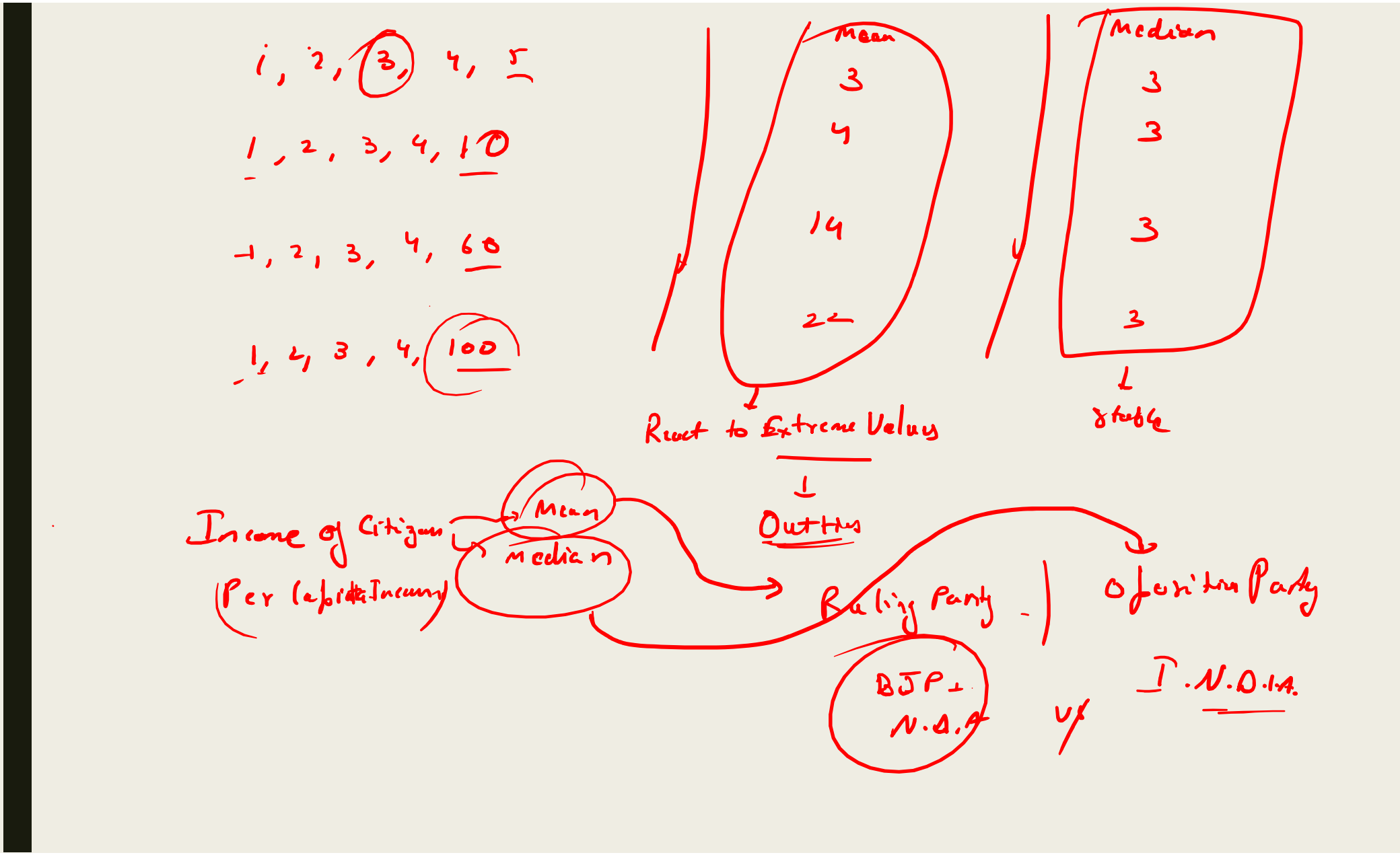
→ Median = $\frac{N+1}{2}^{th}$ term

$N = 51$,

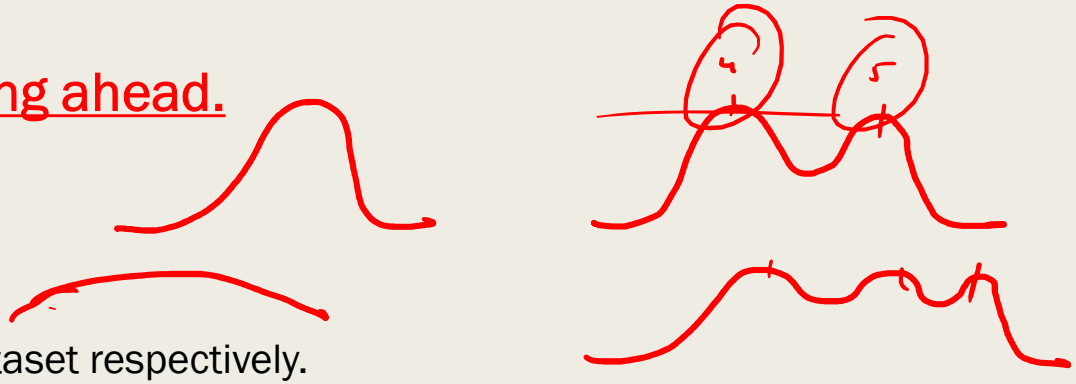
$\frac{51+1}{2}$ = 26^{th} term

↓

median


1, 2, 3, 4, 5
1, 2, 3, 4, 10
1, 2, 3, 4, 60
1, 4, 3, 4, 100
Mean
3
4
14
22
Median
3
3
3
3
React to Extreme Values
Stable
Income of Citizen
Per Capita Income
Mean
Median
Outliers
Ruling Party
Opposition Party
BJP + N.D.A.
I.N.D.I.A.

Let's build some concepts before going ahead.



1. What is Minimum and Maximum value?

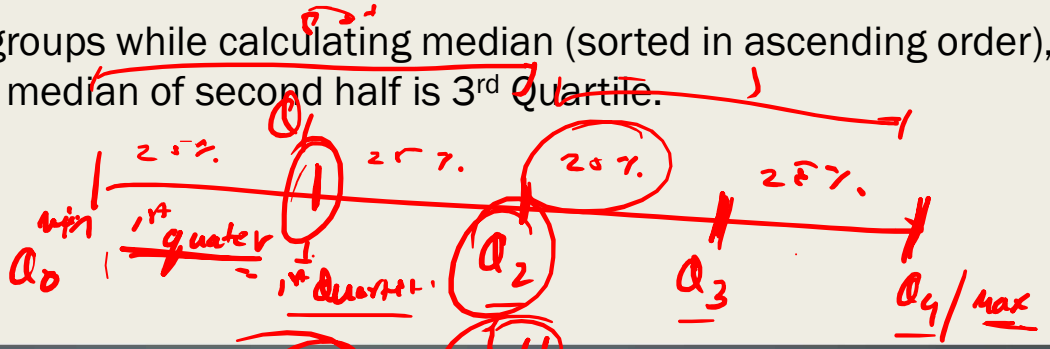
It is the minimum and Maximum values of the dataset respectively.

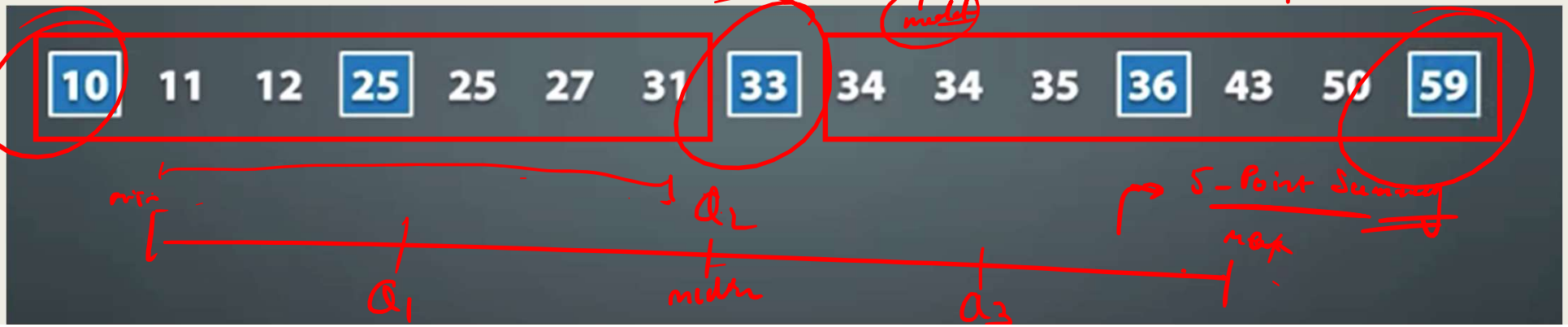
2. What is 1st and 3rd Quartile? – Also called the lower and upper quartile respectively.

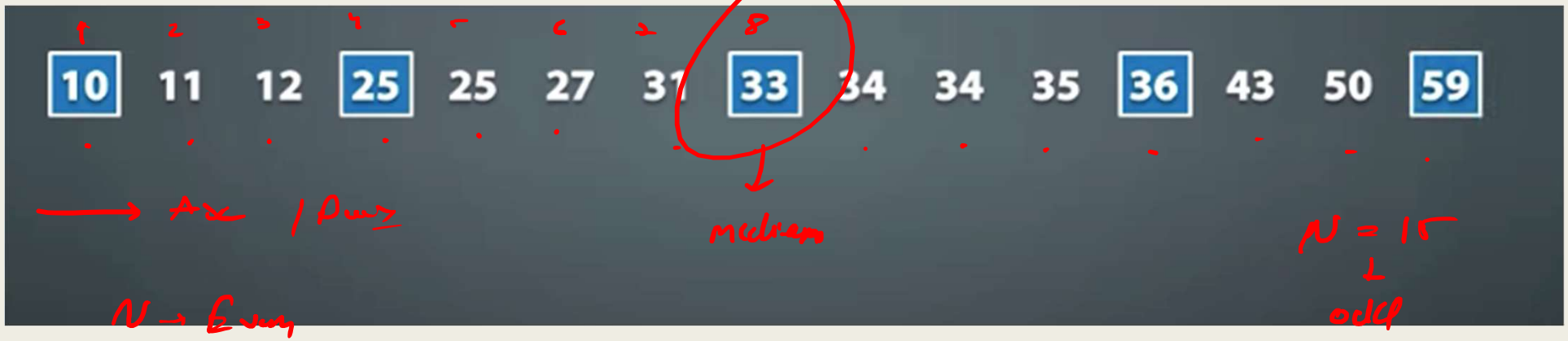
When we divide the dataset into two groups while calculating median (sorted in ascending order), then the median of first half is 1st Quartile and median of second half is 3rd Quartile.

3. Then where is the 2nd Quartile?

Your median is the 2nd Quartile ;-)







→ Asc / Desc

median

N → Even

N = 15

odd

Median = $\frac{N+1}{2}^{th}$ term

$= \frac{15+1}{2} = 8^{th}$ term

$Avg\left(\frac{N}{2}, \frac{N}{2}+1\right)^{th}$ term

$Avg\left(\frac{14}{2}, \frac{14}{2}+1\right)^{th}$

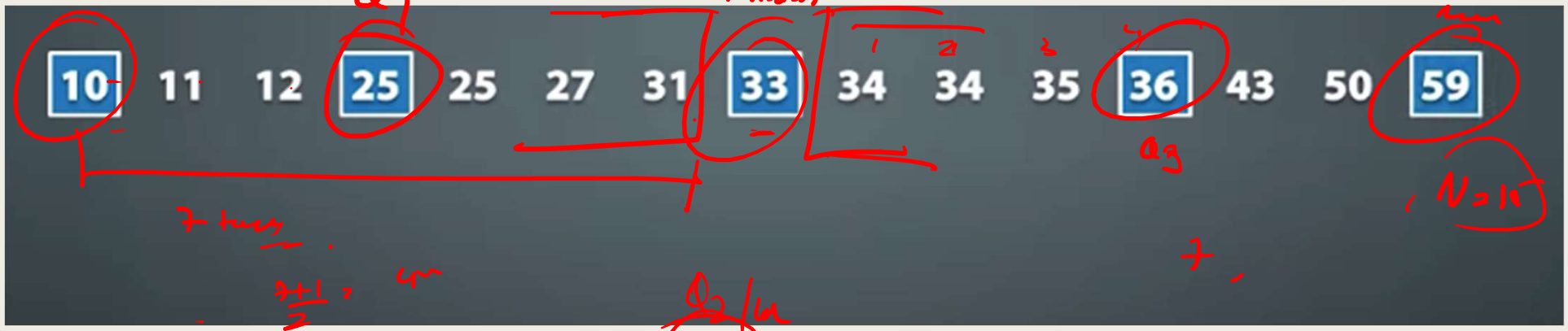
$Avg\left(7, 8\right)^{th}$

59, 50, 43, 36, 35, 34, 33, 33, 31, 27, 25, 25, 12, 11

$Avg\left(34, 36\right) = \frac{33+34}{2} = \frac{67}{2} = 33.5$

median

N = 14



Q0 Q1 Q2/M Q3 Q4

10, 11, 12, 25, 25, 27, 31, 33, 34, 34, 35, 36, 43, 50

1 2 3 4 5 6 7 8 9

$\frac{31+33}{2} = \frac{64}{2} = 32$

N = 14

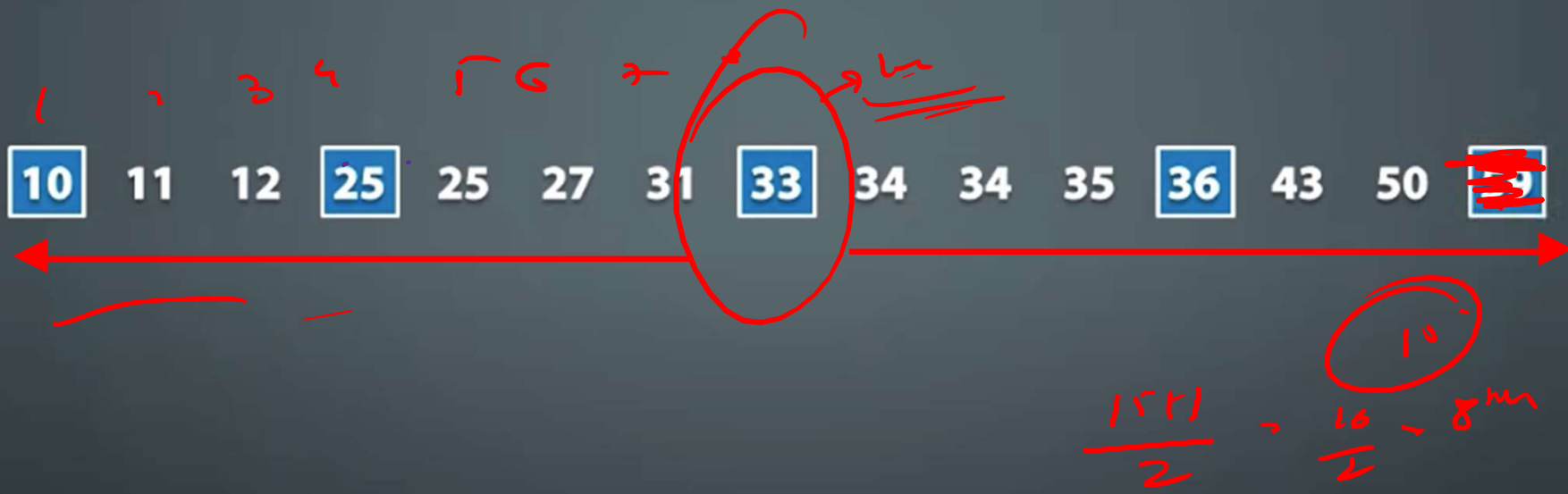
Avg (31, 33)

Avg = (31, 33)

$\frac{31+33}{2} \rightarrow 32$

FIVE NUMBER SUMMARY

MINIMUM	1ST QUARTILE	MEDIAN	3RD QUARTILE	MAXIMUM
10	25	33	36	59



Asc → min, Q_1, Q_2, Q_3, max

Q. Given is the ages of people registered for a webinar, calculate the 5 point summary (5 number summary) of the ages of the participants?

Desc → max, Q_3, Q_2, Q_1, min

19, 26, 25, 37, 32, 28, 22, 23, 29, 34, 39, 31

$\frac{N+1}{2}$ th

$\frac{6}{2}$ (3)

$\frac{6}{2}+1 = 3+1 = 4$

19, 22, 23, 25, 26, 28, 29, 31, 32, 34, 37, 39

24 — 28.5 — 33

6 — 6

Avg (3, 4)th term

$\frac{25+29}{2} = 24$

$\frac{32+34}{2} = 33$

N = 12

Even

$\frac{12}{2} = 6$

$\frac{12}{2}+1 = 6+1 = 7$

Avg $\left(\frac{N}{2}, \frac{N}{2}+1\right)^{th}$ term

Median

Median = Avg $(6^{th}, 7^{th})$ term

⇒ Avg (28, 29)

⇒ $\frac{28+29}{2} = 28.5$

20 Y.

19 — 24 — 28.5 — 33 — 39

min ⇒ 19

Q_1 = 24

Q_2/Median ⇒ 28.5

Q_3 = 33

max = 39