Data and Statistics

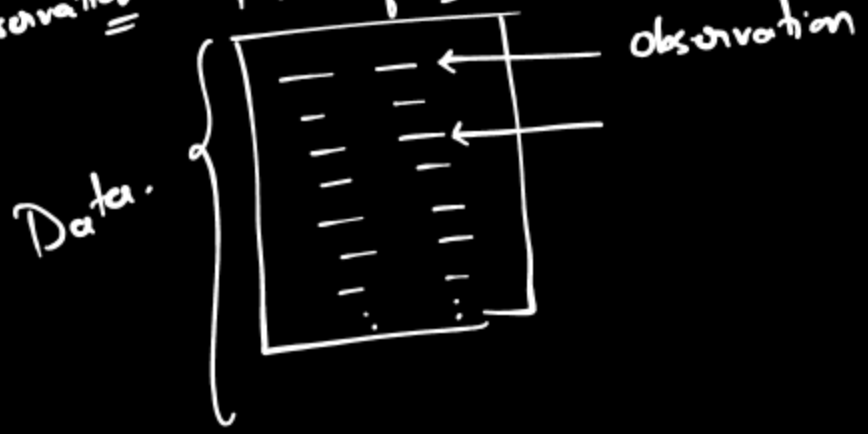
Thanks

Data :
- → profit or loss of a company.
- → Population of a country.
- → monthly wages earned by workers in a factory.
- → Marks obtained by students in a class in certain exam in school

Data: It means information or set of facts presented in numerical & figures.

each fact is called observation.

Marks of 50 students => Data.



"Statistics": It is the science which deals with collection, analysis and interpretation of numerical data. (presentation.)

In singular form, Statistics is a subject.

Fundamental Characters of Data:

- Numerical facts alone constitutes data.
- Qualitative characters like intelligence, poverty, etc. cannot be measured numerically, hence they do not form data.
- Data are aggregate (collection) of facts. Single observation does not form data.

Limitations of Statistics

- Statistics deals groups and does not study individual observations.
- Statistics is not useful for qualitative studies like honesty, poverty, intelligence.
- Statistical laws are not exact.
  - [it represents average values]

Data (collection of numerical facts)

(I) Primary Data

- The data collected by the investigator with a definite plan in mind.
- More reliable and relevant

(II) Secondary Data.

- The data collected by someone, other than the investigator.
  - Not so reliable and relevant.
  - should be carefully used.

Investigator

Raw or Ungrouped data

$\Rightarrow$ The data obtained in original form is raw data or ungrouped data.

Grouped Data: When we condense data into classes or groups, then it becomes grouped data.

Array :- An arrangement of raw numerical data in ascending or descending order is called an array (Arrayed data)

Presentation of Data

→ Putting data in

- → tables ✓ (tabular form)
- → class / group.
- → Charts / graphs.
  - → Scatter plot
  - → Histograms.

Frequency of an observation:

The number of times an observation occurs in our a given dataset is called its frequency.

eg: 4, 5, 2, 3, 9, 4, 11, 2, 7, 5 Dataset

10 observation.

Q. Following data gives the number of children in 20 families.

[4, 3, 2, 4, 2, 2, 1, 3, 3, 2, 3, 3, 2, 1, 1, 4, 3, 2, 1, 1.]

Make an array of the above data and construct a frequency table.

Array:

1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 5

(5) (6) (4) (3) (2)

Header

No. of children	No. of families (frequency)
1	5
2	6
3	4
4	3
5	2

Table

This table is called frequency table aka frequency distribution table

Types of frequency Distribution

0 – 10	
10 – 20	
20 – 30	

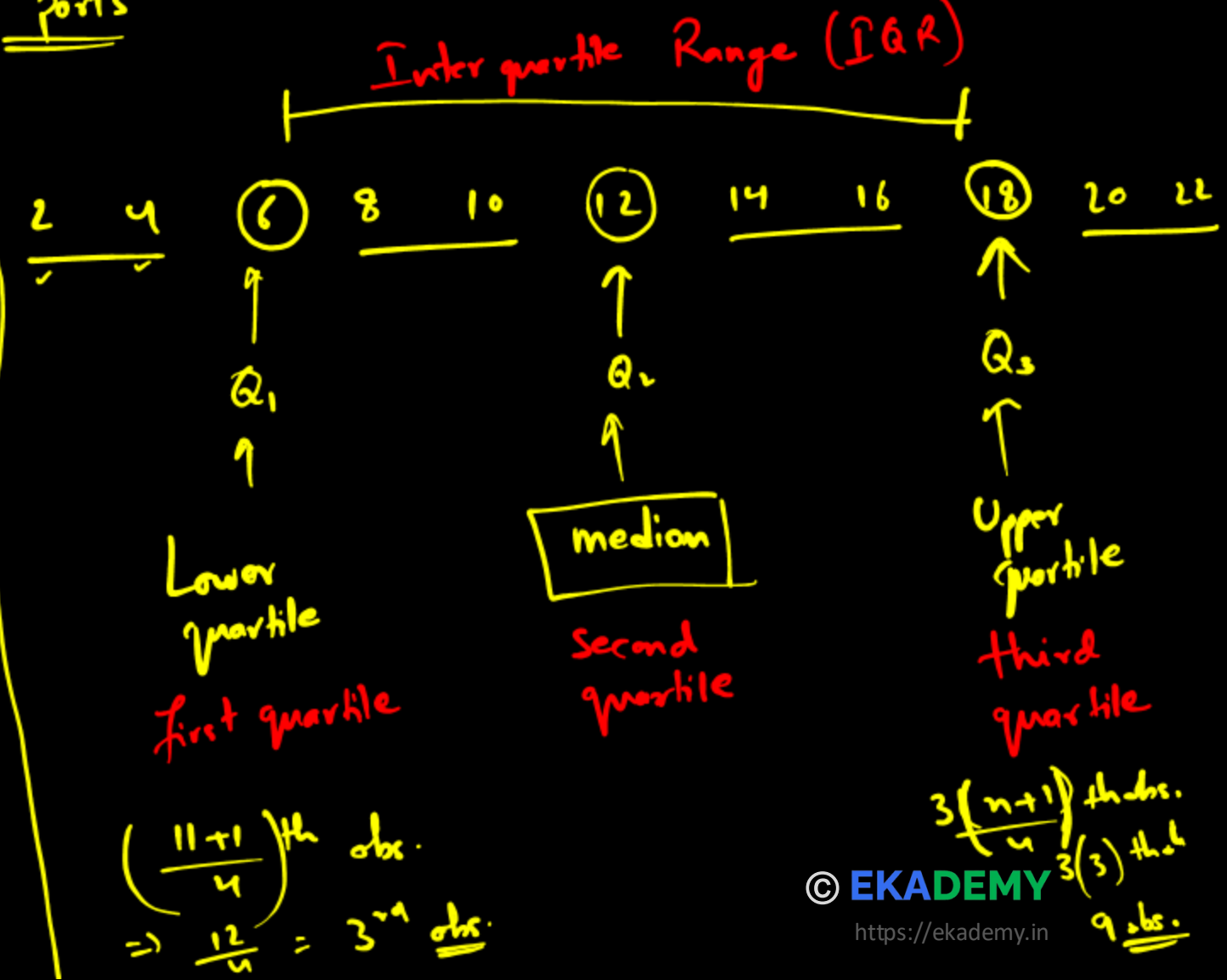
Quartiles

↳ Values in a given data set which divides the whole data set into four equal parts

$\Rightarrow$

Q_1 (lower quartile) $\begin{cases} = \left(\frac{n+1}{4}\right)^{th} \text{ observation} & n \Rightarrow \text{odd} \\ = \left(\frac{n}{4}\right)^{th} \text{obsand} \ (n \Rightarrow \text{even}) \end{cases}$

Q_3 (Upper Quartile) $= \begin{cases} = 3\left(\frac{n+1}{4}\right)^{th} \text{obs.} & n \Rightarrow \text{odd} \\ = 3\left(\frac{n}{4}\right)^{th} \text{obs.} & n \Rightarrow \text{even.} \end{cases}$



Inter Quartile Range (IQR) = $Q_3 - Q_1$

Semi interquartile Range = $\frac{Q_3 - Q_1}{2}$

Q. In a class test, marks scored by 11 students are:

13, 17, 20, 5, 3, 19, 7, 6, 11, 15, 17.

Find: (i) median, (ii) Lower quartile (iii) Upper Quartile (iv) Inter Quartile Range.

11

13

11

6

11

17

11

11

Mode

↳ The observation (value) that occurs most frequently in the given data set.

* Data set may or may not have mode.

for eg: the data 2, 5, 0, 7, 4, 6 has no mode.

* Data set may have one or more than one mode.

↓

Multi-modal data.

eg1: 2, 5, 0, 7, 3, 5, 9, 11.

mode = 5

eg2: 2, 5, 0, 3, 5, 7, 6, 3, 8.

mode: 3 and 5

Empirical Relation between mean, median and mode.

$$\text{Mode} = 3(\text{median}) - 2(\text{mean})$$

Q.

Heights (in cm)	151	152	153	154	155	156	157
No. of students (frequency)	6	4	11	9	16	12	2

$= 60$

Find : (i) mean (ii) median (iii) mode (iv) Q_1 (v) Q_3 (vi) IQR (vii) Verify empirical relation between mean, median and mode.

H.W.

$$\text{Mean} = \frac{(6\times151)+(4\times152)+(11\times153)+(9\times154)+(16\times155)+(12\times156)+(2\times157)}{60}$$

$$= \frac{906+608+1683+1386+2480+1872+314}{60}$$

$$= \frac{9249}{60}$$

mean. = 154.15 cm. (mean height)

Central tendency

→ Central value / position in the given data.

3 methods to find central value

- Mean (arithmetic mean)
- Median
- Mode

Mean

"The average of a given set of numbers is called mean or arithmetic mean.

data

Example: Given nos. ⇒ 2, 4, 6, 8, 10

$$\text{mean} = \frac{2+4+6+8+10}{5} = \frac{30}{5}$$

mean = 6

$\bar{x} = 6$

Symbol for mean is $\bar{x}$

Formula

Data : $\underbrace{x_1, x_2, x_3, \ldots, x_n}_{n - \text{observation}}$

$$\text{mean} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

$$\bar{x} = \frac{\text{Sum of observations}}{\text{number of observations}}$$

Q. Find the mean of first six multiples of 3.

Sol: first six multiples of 3 are: 3, 6, 9, 12, 15, 18

$$\text{mean} = \frac{3+6+9+12+15+18}{6}$$

$$= \frac{63}{6}$$

$$\boxed{\text{mean} = 10.5}$$

→ may or may not be part of dataset.

→ mean can be fractional.

Q. Heights of five basketball players are 153 cm, 140 cm, 148 cm, 150 cm, 154 cm. Find the mean height of the 5 player

$\Rightarrow$ 149 cm

Q. If the mean of 6, 8, 9, x, 13 is 10. Find the value of x.

Sol:

$$\text{mean} = 10$$

$$\text{mean} = \frac{6+8+9+x+13}{5}$$

$$\frac{6+8+9+x+13}{5} = 10$$

$$6+8+9+x+13 = 50$$

$$36+x = 50$$

$$x = 50-36$$

$$\boxed{x = 14}$$

Q. If the mean of the observations $x, x+3, x+5, x+7, x+10$ is 9. Find the mean of last three observations.

Sol:

$$\text{mean} = 9$$

$$\frac{x + x+3 + x+5 + x+7 + x+10}{5} = 9$$

$$\frac{5x + 25}{5} = 9$$

$$\frac{5(x+5)}{5} = 9$$

$$x + 5 = 9$$

$$x = 9 - 5$$

$$\boxed{x = 4}$$

$\therefore$ The last three observations will be, $4+5, 4+7, 4+10$

$= 9, 11, 14$

mean of $9, 11, 14$

$$\text{mean} = \frac{9 + 11 + 14}{3}$$

$$= \frac{34}{3} = 11.33333\ldots$$

$$= 11.\bar{3}$$

$11\frac{1}{3}$

Q. The mean of five numbers is 30. If one number is excluded, their mean becomes 28. Find the excluded number.

Sol:

Mean of 5 obs. is 30

$\therefore$ Sum of these 5 numbers $= 30 \times 5 = 150$

Mean of remaining 4 observation $= 28$

$\therefore$ Sum of these 4 numbers $= 4 \times 28 = 112$

Excluded number $= 150 - 112$

$= 38$

$$\text{Mean} = \frac{\text{Sum of observations}}{\text{no. of observation.}}$$

$$\text{Sum of observations} = \text{Mean} \times \text{no. of obs.}$$

Q. The mean of 40 numbers was found to be 38. Later on, it was detected a number 56 was misread as 36. Find the correct mean of the given numbers.

H.W.

Sol: Calculated mean of 40 nos. = 38

∴ Sum of 40 nos. (calculated) = $40 \times 38 = 1520$

Correct sum = ?

= 1520 − wrong no. + correct no.

= 1520 − 36 + 56

= 1520 + 20

Correct sum = 1540

Correct mean = $\frac{1540}{40}$

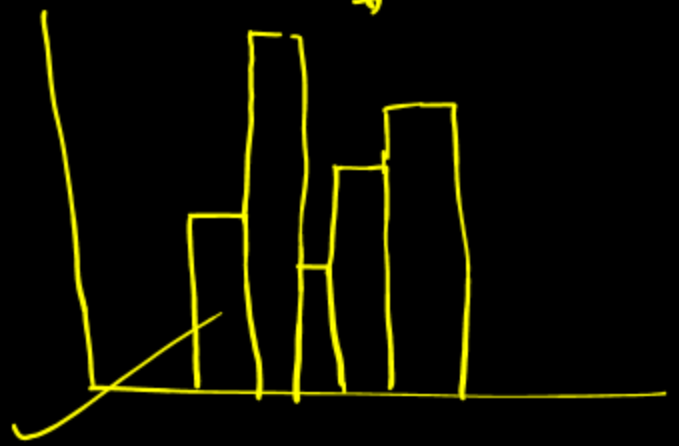
$= \frac{154}{4} = 38.5$

Bar Graph , Histogram , Frequency Polygons.

Ungrouped data (categorical data).

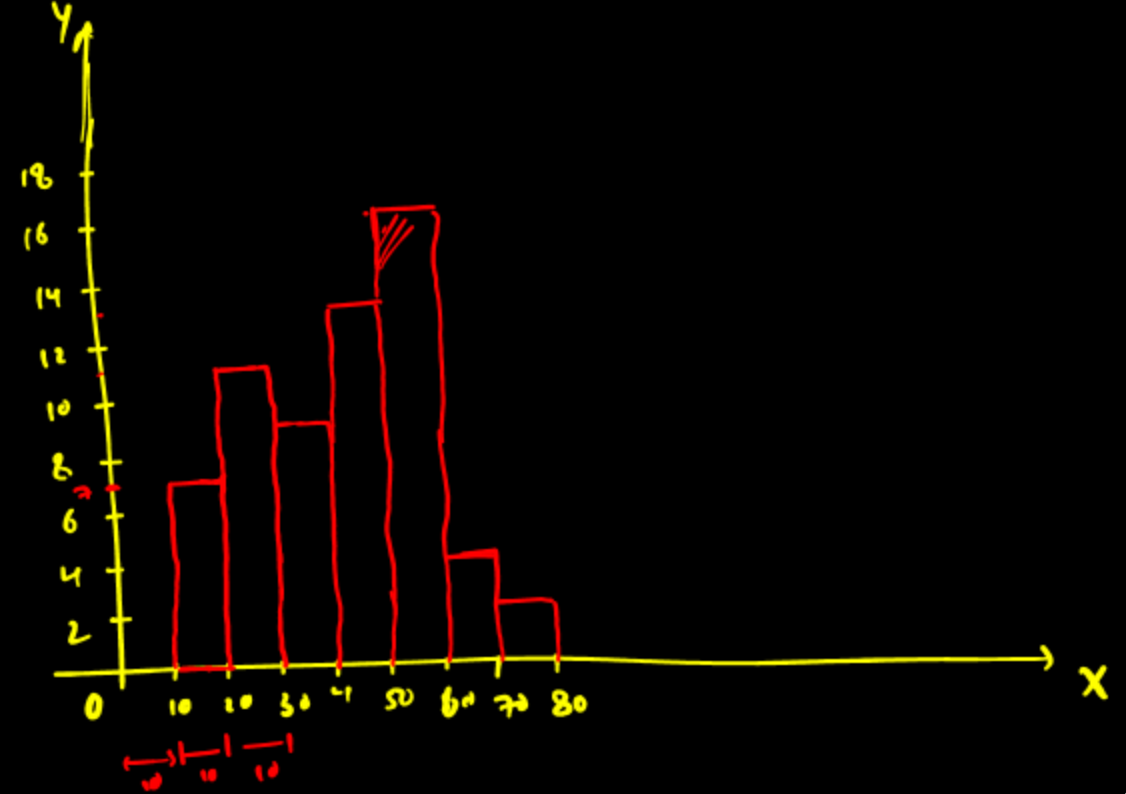
Grouped data.

0 – 10
10 – 20
20 – 30



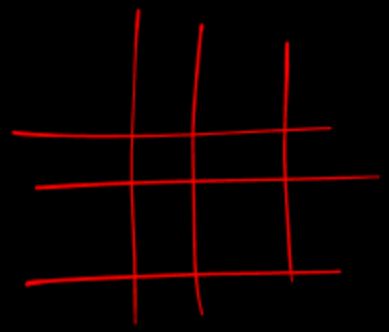
Draw a histograms

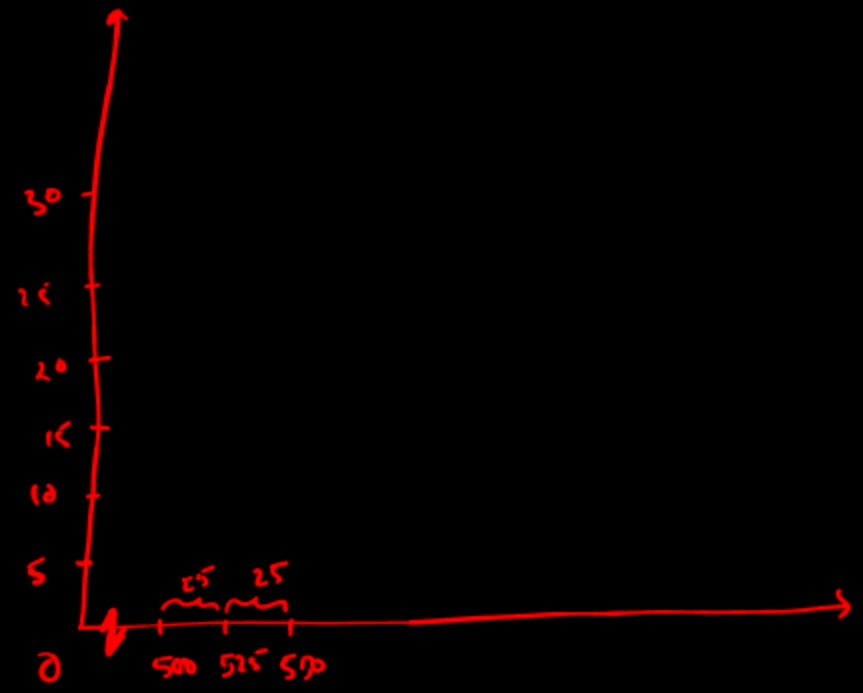
Marks	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80 → X-axis
Number of Students (f)	7	11	9	13	16	4	2 → Y-axis

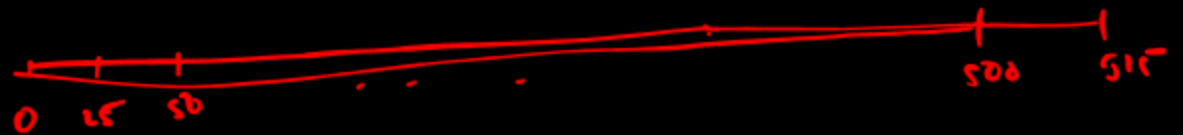


Histogram

Daily wages (₹)	500 – 525	525 – 550	550 – 575	575 – 600	600 – 625
No. of workers (f)	20	15	25	30	10







10 – 19

20 – 29

30 – 39

Histogram from Data in inclusive form.

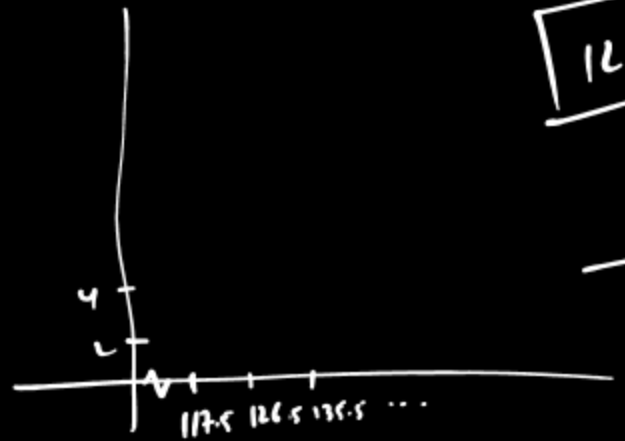
Length (cm)	118 – 126	127 – 135	136 – 144	145 – 153	154 – 162	163 – 171	172 – 180
No. of leaves in the plant (f)	8	10	12	17	7	5	3

exclusive

	Length (cm)	# Leaves
117.5 – 126.5	118 – 126	8
126.5 – 135.5	127 – 135	10
135.5 – 144.5	136 – 144	12
	145 – 153	17
	154 – 162	7
	163 – 171	5
	172 – 180	3

126 – 127

126.5 cm



Adjustment factor = $\frac{\text{Lower limit} - \text{U.L of pre. class}}{2}$

$A.F. = \frac{127 - 126}{2} = \frac{1}{2} = 0.5$

Median

→ Central value (middle observation) in an arrayed data set (ascending | descending order)

eg 1:

⇒ 2, 4, 9, 11, 15, 19, 20

middle observation ↑ median

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

n = 7 (odd) = $\frac{7+1}{2} = \left(\frac{8}{2}\right) = 4^{th}$ observation = 11

eg 2:

⇒ 3, 5, 9, 17, 21, 33, 49, 51, 55, 56 | n = 10 (even no.)

5th, 6th

middle observations

median = mean of middle observation = $\frac{21+33}{2}$

$= \frac{54}{2} = 27$

If there are n observations $x_1, x_2, x_3, \cdots, x_n$

(x_1: first observation; x_n: n^{th} observation)

$$\text{median} = \begin{cases} \left(\frac{n+1}{2}\right)^{th} \text{ observation}, & \text{if } n \text{ is odd.} \\ \dfrac{\frac{n}{2}^{th} \text{ observation} + \left(\frac{n}{2}+1\right)^{th} \text{ observation}}{2}, & \text{if } n \text{ is even.} \end{cases}$$

If $n = 10$,

$$\text{median} = \frac{\left(\frac{10}{2}\right)^{th} \text{ observation} + \left(\frac{10}{2}+1\right)^{th} \text{ observation}}{2}$$

$$= \frac{5^{th} \text{ obs.} + 6^{th} \text{ obs.}}{2} = \frac{21+33}{2} = \frac{54}{2} = 27$$

Q. Find the median of the following data set:

(i) 3, 5, 0, 4, 9, 7, 6, 2, 8

(ii) 3, 5, 1, 2, 4, 6, 0, 2, 2, 3.

(i) Arranging the given numbers in ascending order:

0, 2, 3, 4, 5, 6, 7, 8, 9

5th obs.

$n = 9$ (odd)

$$\text{median} = \left(\frac{n+1}{2}\right)\text{th observation.}$$

$$= \left(\frac{9+1}{2}\right)\text{th obs.}$$

$$= 5^{\text{th}} \text{ obs}$$

$$= 5$$

Q. The mean of the following observation is 68. Find x, and hence find the median.

45, 52, 60, x (= 115), 69, 70, 26, 81, 94

Sol^n: Given obs. = _____.

no. of observation (n) = 9 (odd)

mean = 68

By definition of mean,

$$\text{mean} = \frac{\text{sum of obs.}}{\text{total no. of observation.}}$$

$$= \frac{45 + 52 + 60 + x + 69 + 70 + 26 + 81 + 94}{9}$$

$$68 = \frac{497 + x}{9}$$

$$x + 497 = 68 \times 9$$

$$x = 612 - 497$$

$$\boxed{x = 115}$$

For median,

arrange in order:

26, 45, 52, 60, (69), 70, 81, 94, 115

median: $\left(\frac{9+1}{2}\right)^{th}$ observation

$= 5^{th}$ obs.

$= 69$.

Q. Calculate the median of following data.

Weight (kg)	46	48	50	52	53	54	55
no. of students	7	5	8	12	10	2	1

Sol:

Weight (kg)	No. of students (f)	Cummulative frequency:
46	7	7
48	5	12
50	8	20
52	12	32
53	10	42
54	2	44
55	1	45
	n = 45 ✓	

$n = 45$ (odd)

$\text{median} = \left(\frac{n+1}{2}\right)^{th}$ observation.

$= \left(\frac{46}{2}\right)^{th}$ obs

$= 23^{rd}$ obs.

$= 52$

46 46 46 46 48 48 50 50 52

1 2 ..... 7 .... 8 .... 12^{th} 13^{th} ..... 20^{th} 21^{st} 22^{nd}

52 52

23^{rd} .... 32^{nd}

Q.

No. of goals	0	1	2	3	4	5
No. of matches	2	4	7	6	8	3

Find the median of the goals in this distribution.

$n = 30$

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

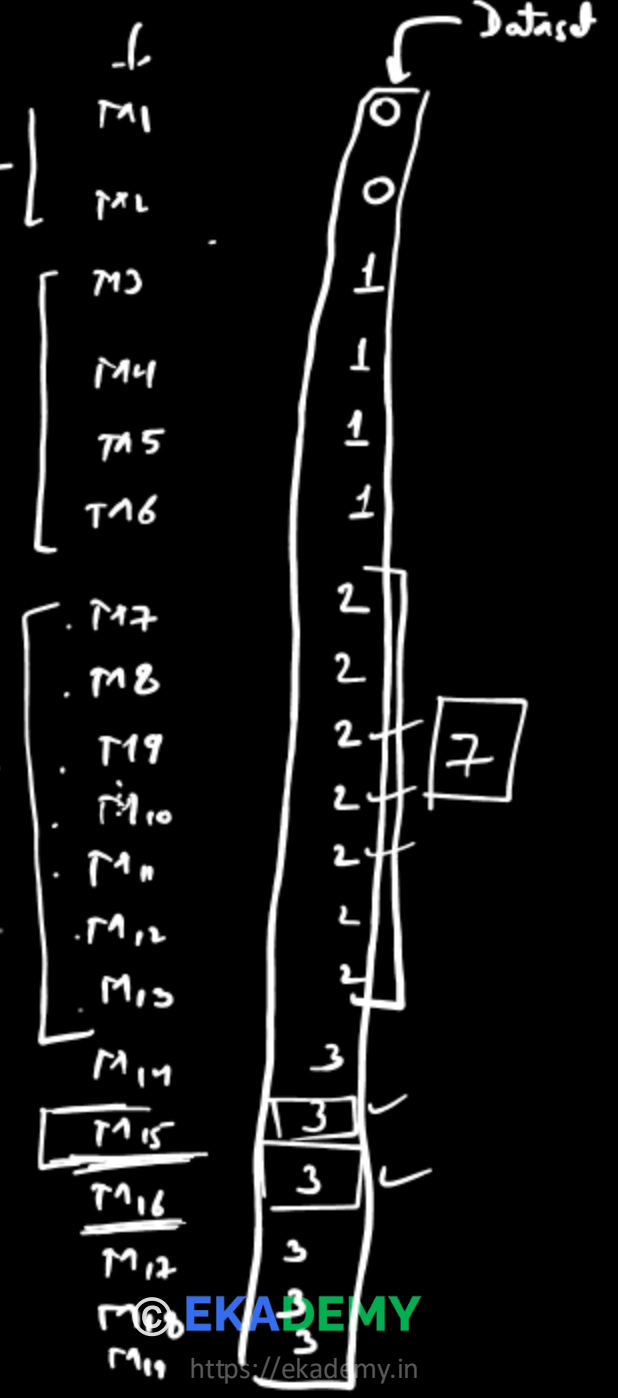
$$= \frac{15^{th} \text{ obs.} + 16^{th} \text{ obs.}}{2}$$

$$= \frac{3+3}{2} = 3$$

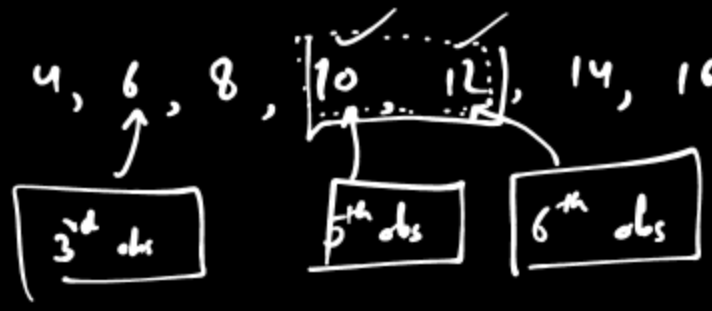
n = even

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

$$\frac{31}{2} = 15.5^{th}$$



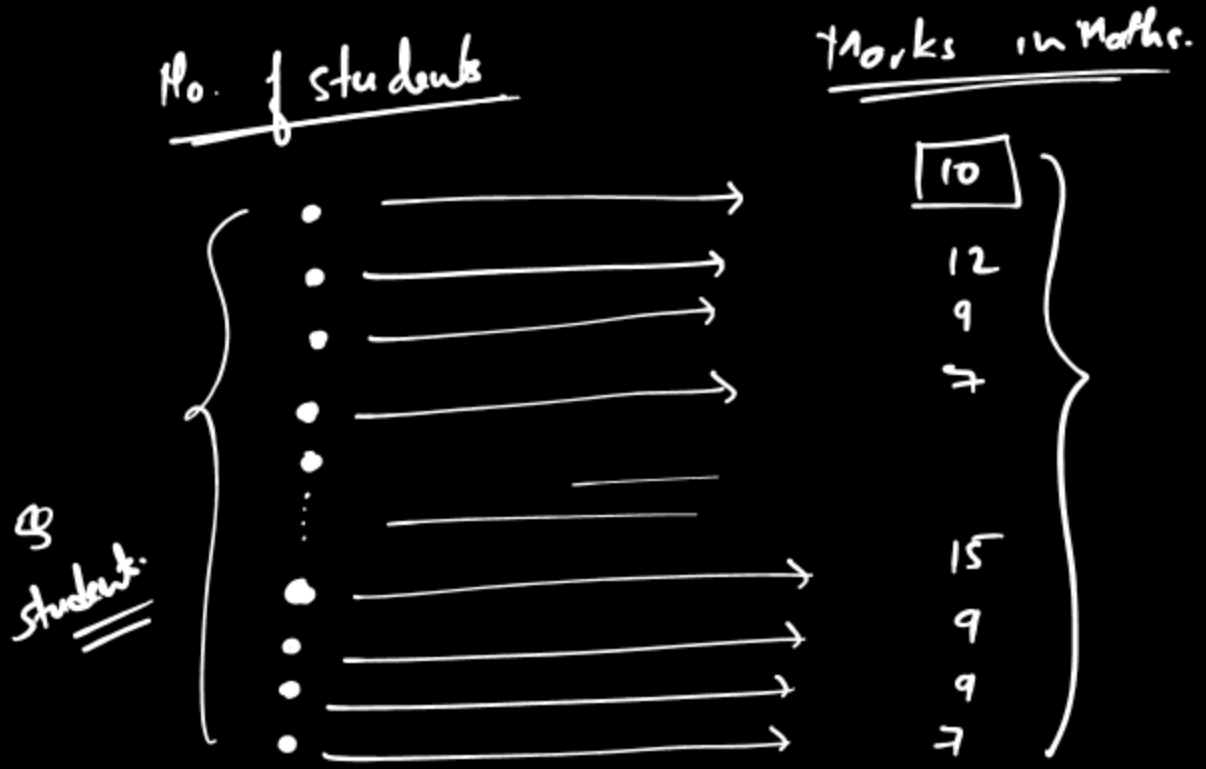
$\Rightarrow$ 2, 4, 6, 8, 10, 12, 14, 16, 18, 20



$$m = \frac{5^{th}\text{ obs} + 6^{th}\text{ obs}}{2}$$

$$\frac{10 + 12}{2}$$

$$= 11$$



Salary of workers : ($)

45, 45, 45, 45, 45, 48, 38, 55, 55, 48, 38, 38, 65, 38, 45,
65, 62, 48, 48, 45, 38, 38, 55, 62, 55, 38,
55, 48, 45, 38, 38, 45, 38, 62, 55, 38, 45,
38, 48, 38, 55, 38, 38, 55, 55, 48.

46 works

Salary ($)	45	48	38	55	65	62
No. of workers	10	7	15	9	2	3

Salary	38	45	48	55	62	65
No. of workers	15	10	7	9	3	2

End of the chapter