IG-Math 0580

Statistics Notes

Suresh Goel
(Director)
Alliance World School
Noida, India.

IG-Maths 0580

Statistics Notes

" Statistics deals with the collection, analysis and interpretation of numerical data"

§ Representation of Data :

1 **Raw Data :** A teacher conducts a test in a class of 20 students and note down the following marks obtained by them, out of maximum marks 10.

7, 8, 5, 9, 3, 10, 7, 9, 4, 8
3, 7, 9, 7, 4, 7, 9, 7, 3, 4

such data are called raw data.

2. **Frequency Distribution:**
It is difficult to make inference if the data are large in number and in raw form.
The above data can expressed as:

Marks Obtained x	Tally Bar	Frequency f
3	|||	3
4	|||	3
5	|	1
7	~~||||~~ |	6
8	||	2
9	||||	4
10	|	1

$n = \Sigma f = 20$

§ Range of the data :

Range = Max. Value of Data − Min. value of Data
$= 10 - 3 = 7$
(In the example above)

§3. Grouped Frequency Distribution:

If the range of the data is large, then we represent the data in the 'grouped frequency distribution'

(i) Discrete Frequence (Inclusive) Distribution:

When the data take only integral values, (and no fractions).
For example: marks obtained, number of workers etc.

Example 1: Form a discrete frequency (inclusive) distribution table for the given raw data;
Given the marks obtained in maths test in a class of 35 students out of max. marks 100.

67, 98, 75, 13, 26, 89, 7, 34, 49, 52, 36, 100,
71, 99, 68, 55, 37, 88, 47, 85, 96, 77, 88, 63
21, 72, 83, 64, 96, 56, 93, 65, 33, 78, 9.

Marks Obtained x	Tally Bars	Frequency f
1 - 10	ll	2
11 - 20	l	1
21 - 30	ll	2
31 - 40	llll	4
41 - 50	ll	2
51 - 60	lll	3
61 - 70	~~llll~~	5
71 - 80	~~llll~~	5
81 - 90	~~llll~~	5
91 - 100	~~llll~~ l	6
		$N = \Sigma f = 35$

Note:
(i) In class 11-20 both the marks 11 and 20 are included.
(ii) Class size $= (20-11)+1$
$= 9+1$
$= 10$ ✓
(iii) for class 11-20
lower limit = 11
upper limit = 20

(iv) for class 11-20; Lower Bound = 10.5 and upper Bound = 20.5

IG-Maths 0580	Statistics	Notes	P-3

§ 3(ii) Continuous Grouped Frequency Distribution:

Continuous data, are, for example, height of plants, weight of students in a class etc.

Example 2: The heights of 25 plants in cm. is given below, represent the data in continuous grouped frequency,

37, 43, 50, 49, 52, 45, 57, 35, 40, 54
36, 44, 52, 35.31, 55.45, 40.91, 36.01,
41.3, 55.5, 51.4, 36, 40.20, 39, 51, 55.

	Height of Plants x	Tally Bars	Frequency f
$35 \leq x < 40$	35 - 40	|||| ||	7
$40 \leq x < 45$	40 - 45	|||| |	6
$45 \leq x < 50$	45 - 50	||	2
$50 \leq x < 55$	50 - 55	|||| |	6
$55 \leq x < 60$	55 - 60	||||	4

$n = \Sigma f = 25$

Note: (i) Value 40 is included in $40 \leq x < 45$ or 40-45
But not in $35 \leq x < 40$ or (35-40)
In continuous frequence distribution upper class limit (or upper class boundry) is not included in the class.

(ii) Class size of $\begin{cases} 35\text{-}40 \text{ is } 40-35=5 \\ 40\text{-}45 \text{ is } 45-40=5 \end{cases}$

(iii) In continuous group distribution.
Classes

$35 \leq x < 40$		35-40	may be written in short as	35 —
$40 \leq x < 45$		40-45		40 —
$45 \leq x < 50$	or	45-50		45 —
$50 \leq x < 55$		50-55		50 —
$55 \leq x < 60$		55-60		55 —

Statistics

§ Statistical Diagrams:

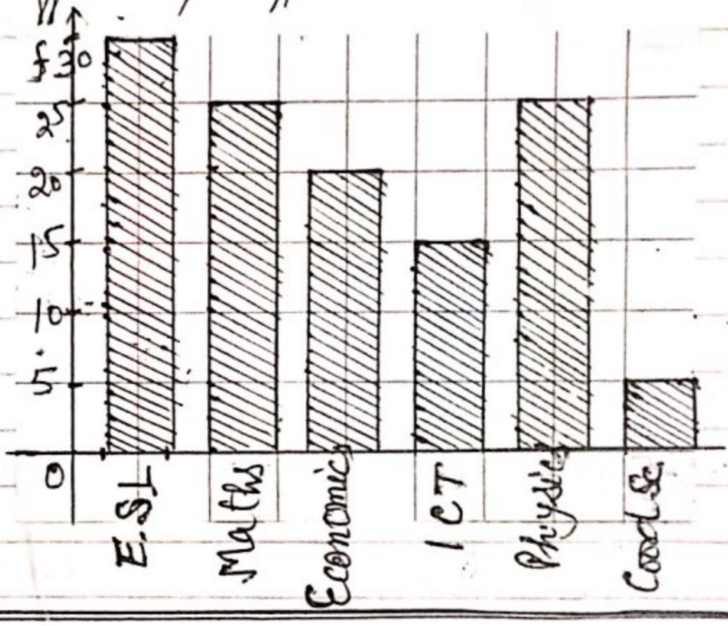
1. Bar Charts:

Discrete data can be displayed by bar charts.

Example 3: The following frequency table show the number of student of IGCSE offering different subjects.

Subject	frequency
ESL	30
Maths	25
Economics	20
ICT	15
Physics	25
Coord Sc.	5

Draw a Bar Chart.



2. Pie Chart:

Data are represented by the sectors of circle. The size of a sector is in direct proportion to the frequency of data.

$$\text{Sector angle} = \frac{f}{\text{Sum of frequency}} \times 360^\circ$$

Example 4:

Country	No of holidays	Sector Angle θ
Thailand	24	$\frac{24}{72} \times 360 = 120^\circ$
Hong Kong	6	$\frac{6}{72} \times 360 = 30^\circ$
Singapore	12	$\frac{12}{72} \times 360 = 60$
Malaysia.	30	$\frac{30}{72} \times 360^\circ = 150^\circ$

Total No of holidays = 72

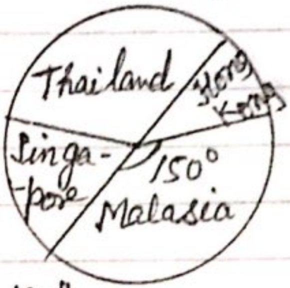
Sector angle,

$$\theta = \frac{\text{No of holidays}}{72} \times 360^\circ$$

Statistics

& Pie Charts.

Example 5: A travel brochure has 72 holidays in four different countries. A pie Chart shows this information.



(a) There are 24 holidays in Thailand. Show that the sector angle for Thailand is 120°. ---[2]

(b) The sector angle for Malaysia is 150°. The sector angle for Singapore is twice the sector angle for Hong Kong. ---[3]
Calculate the number of holidays in Hong Kong. S-14/21/Q17

Solution

(a) Total no. of holidays = 72
Holidays in Thailand = 24

$$\text{Sector angle} = \frac{\text{Holidays in Thailand}}{\text{Total Holidays}} \times 360$$

$$= \frac{24}{72} \times 360 = \underline{120°} \checkmark$$

(b) Let the sector angle for Hong Kong $= x°$

$\therefore$ sector angle for Singapore $= 2x°$

$$\text{Total} = x° + 2x° + \underset{\text{Thai}}{120°} + \underset{\text{Malaysia}}{150°} = 360°$$

$$\therefore 3x + 270 = 360°$$

$$3x = 90°$$

$$x = 30°$$

$\therefore$ No. of holidays in Hong Kong

$$= \frac{30°}{360°} \times 72$$

$$= 6° \checkmark$$

Example 6: Michelle sell ice cream. The table shows how of different flavours she sells in one hour.

Flavour	Vanilla	Strawberry	Chocolate	Mango
Number Sold	6	8	9	7

Michelle wants to show this information in a piechart. S-14/23/Q2
Calculate the sector angle for Mango. --- [2]

Solution:
$$\text{Sector angle for Mango} = \frac{\text{No of Mango ice cream}}{\text{Total no of ice cream}} \times 360° = \frac{7}{30} \times 360 = \underline{84°} \checkmark$$

Statistics

Statistical Diagrams:

§ Histogram:

A histogram is a graphical representation of a frequency distribution in the form of rectangles with class intervals as bases and heights proportional to the corresponding frequencies such that there is **no gap** between any two successive rectangles.

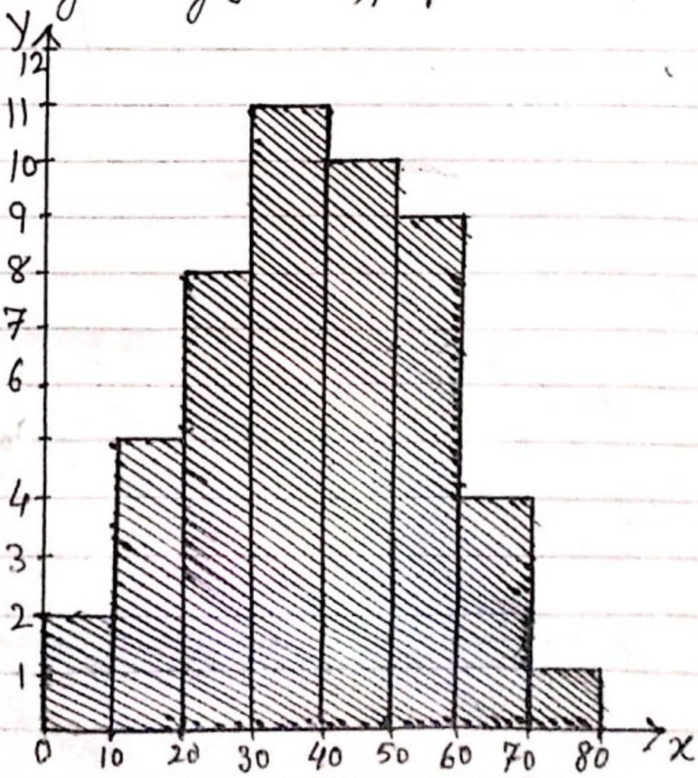
Case I: A continuous grouped frequency distribution with equal class-intervals.

Example 7: The table gives the marks scored by 50 students in a maths test.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No of Students	2	5	8	11	10	9	4	1

Represent this data in the form of a histogram.

Solution: Class size in all the intervals (classes) is same, so we can represent the classes on x-axis with equal width and frequency along y-axis, proportional to the the frequencies as the height of the histograms.



Statistics

Statistical diagrams

Histogram :

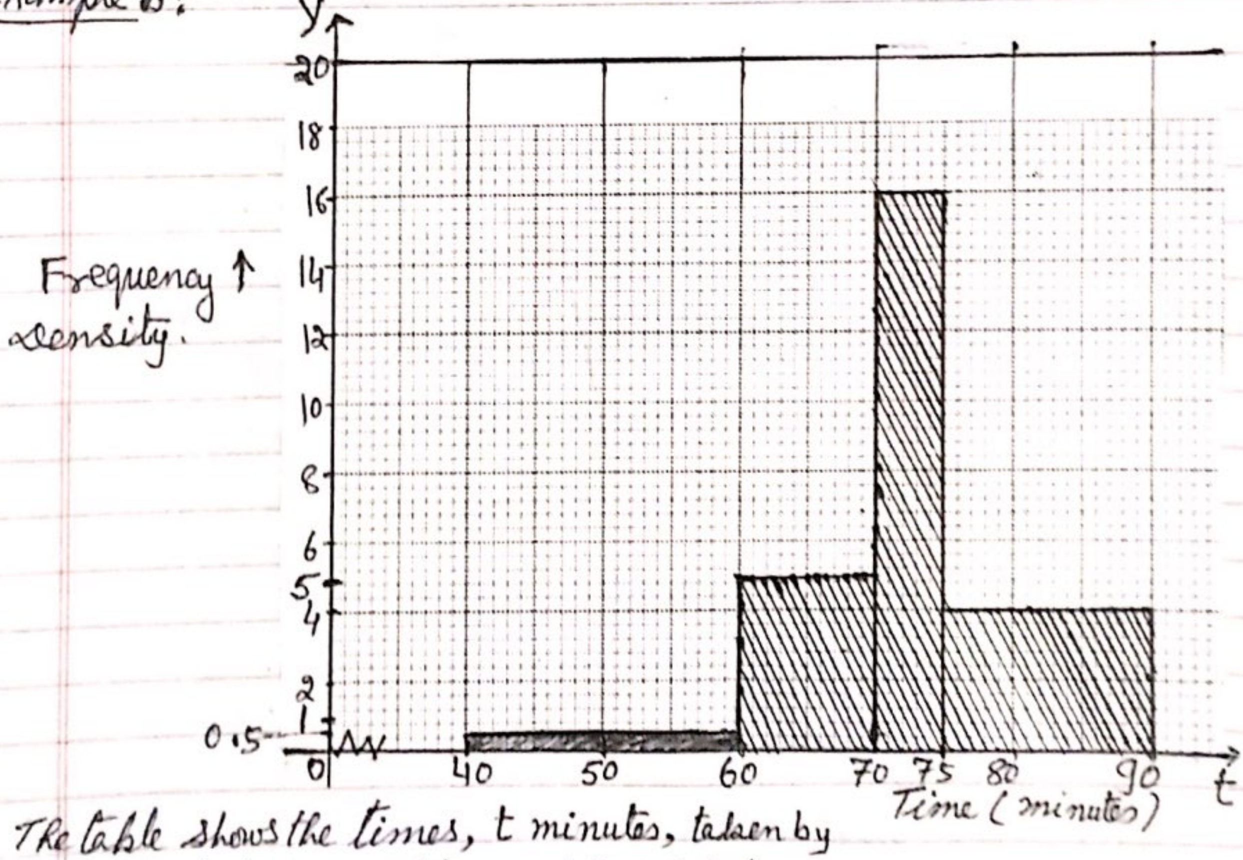
Case II. A continuous frequency distribution with unequal class-sizes of different classes.

Along x- axis the width of the histogram is proportional to the class-size.

Along y - axis we take height of the histogram proportional to the frequency density the class.

$$\text{Frequency density} = \frac{\text{frequency}}{\text{class width}}.$$

Example 8:



The table shows the times, t minutes, taken by 200 students to complete an IGCSE paper.

Time (t min)	$40 < t \leq 60$	$60 < t \leq 70$	$70 < t \leq 75$	$75 < t \leq 90$
Frequency	10	50	80	60

Solution

Frequency Density	$\frac{10}{20} = 0.5$	$\frac{50}{10} = 5$	$\frac{80}{5} = 16$	$\frac{60}{15} = 4$

On the grid, draw a histogram to show the information in the table.

S-15/43/Q4(b) --- [4]

Statistical Diagrams.

Histogram:

§ Case III A discrete grouped frequency distribution is given.

Example 9: 100 students appear for an A-Level Physics exam, their marks (as percentages) are given in the table below, Draw a Histogram.

Marks %	Frequency
41-50	15
51-60	10
61-80	40
81-85	25
86-100	10

$n = \Sigma f =$

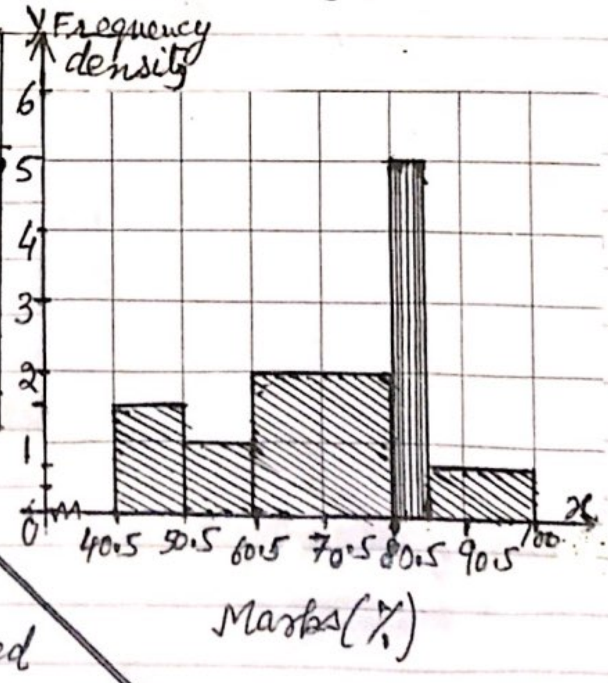
First we should change class limits to class boundaries. For that note the difference between the lower limit of a class (51-60) and the upper limit of the preceding class (41-50)

$\therefore$ difference $= 51 - 50 = 1$

Now take half the diff $= \frac{1}{2} = 0.5$

$\therefore$ to convert the given distribution into continuous, subtract 0.5 from the lower limit and add 0.5 to the upper limit. Also the class-size of each class is not same so we shall find the Frequency density of each class.

Class Boundaries (Marks %)	Frequency f	Frequency density $= \frac{f}{\text{class size}}$
40.5-50.5	15	$\frac{15}{10} = 1.5$
50.5-60.5	10	$\frac{10}{10} = 1$
60.5-80.5	40	$\frac{40}{20} = 2$
80.5-85.5	25	$\frac{25}{5} = 5$
85.5-100	10	$\frac{10}{15.5} = 0.65$

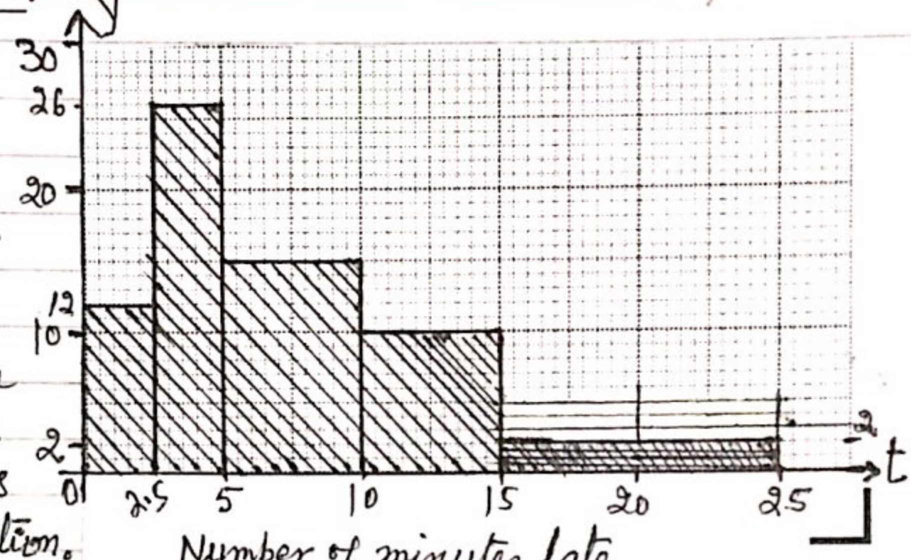


Note: upper class-boundary of the last class is not 100.5 as the marks percentage can't exceed 100%.

Statistical Diagrams

Histogram:

Example 10:



Deborah records the number of minutes late, t, for trains arriving at a station. The histogram shows the information.

(a) Find the number of trains that Deborah recorded. ---[2]

(b) Calculate the percentage of the trains recorded that arrived more than 10 minutes late. [S-16/22/Q20] ---[2]

Solution:

(a) $\text{Frequency density} = \dfrac{\text{frequency}}{\text{Class size}}$

$\therefore \text{frequency} = \text{Frequency density} \times \text{class size}.$

$\therefore$ Total number of trains $= \Sigma f = 12 \times 2.5 + 26 \times 2.5 + 15 \times 5 + 10 \times 5 + 2 \times 10$

$= 30 + 65 + 75 + 50 + 20$

$= \underline{240}$ ✓

(b) No. of trains more than 10 minutes late $= 50 + 20 = 70$

$\therefore$ Percentage of trains which are more than 10 minutes late $= \dfrac{70}{240} = 29.166$

or $= \underline{29.2\%}$

Statistical Diagrams.

Histogram:

Example 11: The table shows information about the time, t minutes, taken for each of 150 girls to complete an essay:

Time (t min)	$60<t\leq 65$	$65<t\leq 70$	$70<t\leq 80$	$80<t<100$	$100<t\leq 150$
Frequency	10	26	34	58	22

The information in the frequence table is shown in a histogram, the height of the block for $60<t\leq 65$ interval is 5 cm.

Complete the table. [W-17/43/Q4]....[3]

Time (t-min)	$60<t\leq 65$	$65<t\leq 70$	$70<t\leq 80$	$80<t\leq 100$	$100<t\leq 150$
Heigh of Block (c.m)	5				

Solution: The height of a block in histogram is proportional to the frequency density.

$$\text{Frequency density} = \frac{\text{frequency}}{\text{Class size}}$$

Now the frequency density for class $60<t\leq 65 = \frac{10}{5} = 2$ unit

But height given $= 5$ cm.

Scale → $\therefore$ 1 unit of f.d $= \frac{5}{2} = 2.5$ cm

Now Again

Time (t min)	$60<t\leq 65$	$65<t\leq 70$	$70<t\leq 80$	$80<t\leq 100$	$100<t\leq 150$
Frequency	10	26	34	58	22
Fre. density	$\frac{10}{5} = 2$	$\frac{26}{5} = 5.2$	$\frac{34}{10} = 3.4$	$\frac{58}{20} = 2.9$	$\frac{22}{50} = 0.44$
Height of Block = f.d × 2.5	$2\times 2.5 = 5$ cm Given	$5.2\times 2.5 = 13$	$3.4\times 2.5 = 8.5$	$2.9\times 2.5 = 7.25$	$0.44\times 2.5 = 1.1$

$\therefore$ Heights of the remaing blocks are: 13, 8.5, 7.25, 1.1

Statistical Diagrams

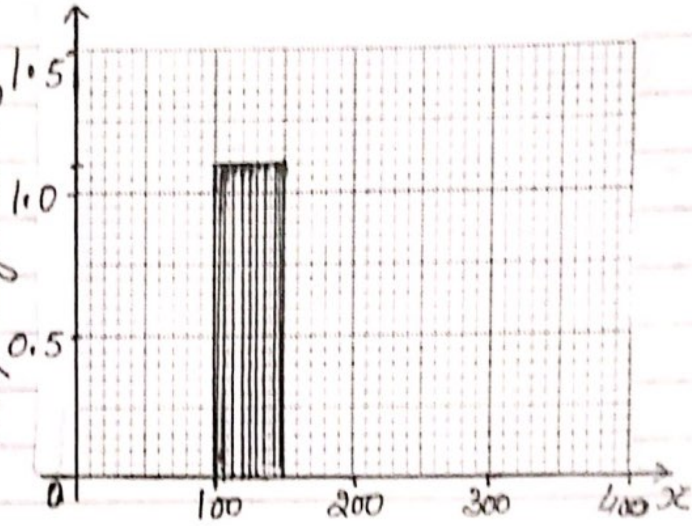
Histogram:

Example 12:

200 students estimate the capacity, x millilitres, of a cup. The results are shown by the frequency table:

Complete the Histogram. ... [4]

[S-17/42/Q3(a)(ii)]



Capacity (x ml)	$0 < x \leq 100$	$100 < x \leq 150$	$150 < x \leq 200$	$200 < x \leq 250$	$250 < x \leq 400$
Frequency	20	55	66	35	24
Fr. density	0.2	1.1	1.32	0.7	0.16

$$\text{Frequency density} = \frac{\text{Frequency}}{\text{Class size}}$$

For the Block ($100 < x \leq 150$) $\quad F.d = \frac{55}{50} = 1.1$ Unit ✓

(Height for) On the diagram 1.1 Unit = 22 small div.

$\therefore$ Scale is 1 Unit = 20 small division

(i) for the Class $0 < x \leq 100$; $F.D = \frac{20}{100} = 0.2$ ✓ $\Rightarrow$ height $= 0.2 \times 20 = 4$ div

(ii) for $150 < x \leq 200$; $F.D = \frac{66}{50} = 1.32$ ✓ $\Rightarrow$ height $= 1.32 \times 20 = 26.4$ division

(iii) for $200 < x \leq 250$,

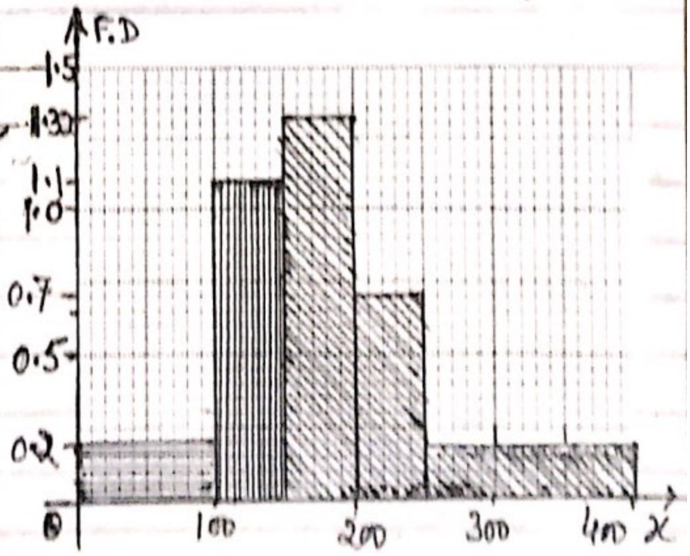
$FD = \frac{35}{50} = 0.7$ ✓

$\Rightarrow$ height $= 0.7 \times 20 = 1.4$ div ✓

(IV) for $250 < x \leq 400$

$FD = \frac{24}{150} = 0.16$ ✓

height $= 0.16 \times 20 = 3.2$ div ✓



§ There are three measures of averages:

(i) Mean (ii) Median (iii) Mode.

§ 1. Mean:

§(i) Mean of Raw data: Given n values $x_1, x_2, x_3, --- x_n$ mean denoted by $\bar{x} = \frac{\Sigma x_i}{n} = \frac{x_1+x_2+x_3 --- x_n}{n}$
and $\boxed{\Sigma x_i = n \cdot \bar{x}}$

Example 13: The scores of five students in class test out of 20 are given as 15, 17, 8, 12, 16, find the mean.

Solution Mean $= \frac{\Sigma x_i}{n} = \frac{15+17+8+12+16}{5} = \frac{68}{5} = \underline{13.6}$ ✓

Example 14: Jim scores the following marks in 8 tests,
7, 8, 8, y, 6, 9, 10, 5, His mean mark is 7.5
Calculate the value of y. --- [2]

Solution Mean marks $= \frac{\Sigma x_i}{n} = \frac{7+8+8+y+6+9+10+5}{8}$

or mean $= \frac{53+y}{8} = 7.5$ given

$\therefore 53+y = 7.5 \times 8$

$y = 60 - 53 = 7$ ✓

$\therefore y = 7$ $\boxed{W-15/22/Q22}$

Example 15:
Amber's mean marks on five tests is 80. Her marks on --- [2]
four of these tests are 68, 81, 74 and 89. $\boxed{W-17/21/Q4}$
Work out her marks on the fifth test.

Solution: Sum of marks on four tests $= 68 + 81 + 74 + 89 = 312$ ✓
and Sum of all five test $\Sigma x_i = n \cdot \bar{x} = 5 \times 80$ (Given Mean $\bar{x} = 80$)
$= 400$ ✓

$\therefore$ Her marks on the fifth test $= 400 - 312$
$= \underline{88}$ ✓

§ 1(ii) Mean of Ungrouped frequency distribution:

Given $\begin{cases} x_i ; & x_1 \quad x_2 \quad x_3 \; ---- \; x_n \\ f_i ; & f_1 \quad f_2 \quad f_3 \qquad\quad f_n \end{cases}$

$$\boxed{\text{Mean } \bar{x} = \frac{\Sigma f_i \cdot x_i}{\Sigma f_i}}$$ (Total number of Values $N = \Sigma f_i$)

Example 16: The table shows information about the numbers of pets owned by 24 students.

Number of pets	0	1	2	3	4	5	6
Frequency	1	2	3	5	7	3	3

(a) Calculate the mean number of pets. ---[3]

(b) Jennifer joins the group of 24 students. When information for Jennifer is added to the table, the new mean is 3.44.
Calculate the number of pets Jennifer has. --[3]

W-15/22/Q22

Solution:

Number of Pets x_i	Frequency f_i	$f_i \cdot x_i$
0	1	0
1	2	2
2	3	6
3	5	15
4	7	28
5	3	15
6	3	18
	$N = \Sigma f_i = 24$	$\Sigma f_i \cdot x_i = 84$

(a) $\text{Mean} = \frac{\Sigma f_i \cdot x_i}{\Sigma f_i}$

$= \frac{84}{24} = \underline{3.5}$ ✓

(b) Now when Jennifer joins.

Total number of students $= 24+1$

$= 25$ ✓

New Mean $= 3.44$

$\therefore$ Sum of Pets $= n \cdot \bar{x}$

for 25 students $= 25 \times 3.44$

$= 86$

and Sum of Pets of 24 students $= 84$

$\therefore$ Number Pets with Jennifer $= 86 - 84$

$= \underline{2}$ ✓

Scanned with CamScanner

Mean

Example 17(a) In a spelling test, marks are given in the table below:
Find the mean.

Marks	0	1	2	3	4	5
Frequency	2	4	5	5	6	8

--- [3]

(b) The table shows the marks gained by some students in English Test.

Marks	52	75	91
Number of Students	x	45	11

--- [3]

The mean marks for these students is 70·3.
Find the value of x.

S-17/43/Q8

Solution:

(a) Mean $= \dfrac{\Sigma x_i \cdot f_i}{\Sigma f_i}$

$= \dfrac{93}{30}$

$= 3.1$

Marks x_i	Frequency f_i	$f_i \times x_i$
0	2	0
1	4	4
2	5	10
3	5	15
4	6	24
5	8	40
	$N = \Sigma f_i = 30$	$\Sigma f_i x_i = 93$

(b) Mean $= \dfrac{\Sigma f_i \cdot x_i}{\Sigma f_i}$

$= \dfrac{76 + 52x}{(56 + x)} = 70.3$ (Given)

Marks x_i	No. of Students f_i	$f_i \cdot x_i$
52	x	$52x$
75	45	3375
91	11	1001

$\Sigma f_i = (56 + x)$; $\Sigma f_i x_i = (4376 + 52x)$

$\Rightarrow 70.3(56 + x) = 4376 + 52x$

$70.3x + 3936.8 = 4376 + 52x$

or $70.3x - 52x = 4376 - 3936.8$

$18.3x = 439.2$

$x = \dfrac{439.2}{18.3} = 24$

$\therefore x = 24$

Mean

§ 1(iii) Mean of Grouped frequency distribution (Estimated Mean):

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$, where x_i is the mid value of interval.

Example 17: The heights, in metres, of 200 trees in a park are measured:

Height (h m)	$2 < h \leq 6$	$6 < h \leq 10$	$10 < h \leq 13$	$13 < h \leq 17$	$17 < h \leq 19$	$19 < h \leq 20$
Frequency	23	47	45	38	32	15

Calculate an estimate of the mean height. [S-13/22/Q20] --[4]

Solution:

for Interval $2 < h \leq 6$

Mid Value $= \frac{2+6}{2}$

$= 4$ ✓

Height (h m)	Frequency f_i	Mid Value x_i	$f_i \cdot x_i$
$2 < h \leq 6$	23	4	$23 \times 4 = 92$
$6 < h \leq 10$	47	8	$47 \times 8 = 376$
$10 < h \leq 13$	45	11.5	$45 \times 11.5 = 517.5$
$13 < h \leq 17$	38	15	$38 \times 15 = 570$
$17 < h \leq 19$	32	18	$32 \times 18 = 576$
$19 < h \leq 20$	15	19.5	$15 \times 19.5 = 292.5$
	$N = \Sigma f_i = 200$		$\Sigma f_i \cdot x_i = 2424$

$$\therefore \text{Mean} = \frac{\Sigma f_i \cdot x_i}{\Sigma f_i}$$

$$= \frac{2424}{200}$$

$$= 12.12$$

$$\therefore \text{Mean} = 12.1$$ ✓

<u>Note</u>: When we find the mean of group data, we assume that all the values of data are at the "mid value" of each class interval, where as the data-values are randomly distributed in each class. That is why, what we are finding by the method above, is called '**Estimated**' mean.

Mean

Mean of Grouped frequency distribution:

Example 18: The table shows information about the time taken by 400 people to complete a race. ---[4]

Time taken (m - minutes)	$45 < m \le 50$	$50 < m \le 60$	$60 < m \le 70$	$70 < m \le 90$	$90 < m \le 100$	$100 < m \le 120$
Frequency	23	64	122	136	26	29

Calculate an estimate of the mean time taken. [M-17/42/Q7(a)]

Solution:

Time Taken (m - minutes)	Frequency f_i	Mid Value of Class Interval x_i	$f_i \cdot x_i$
$45 < m \le 50$	23	$\frac{45+50}{2} = 47.5$	$23 \times 47.5 = 1092.5$
$50 < m \le 60$	64	$\frac{50+60}{2} = 55$	$64 \times 55 = 3520$
$60 < m \le 70$	122	$\frac{60+70}{2} = 65$	7930
$70 < m \le 90$	136	$\frac{70+90}{2} = 80$	10880
$90 < m \le 100$	26	$\frac{90+100}{2} = 95$	2470
$100 < m \le 120$	29	$\frac{100+120}{2} = 110$	3190

$N = \Sigma f_i = 400$ $\qquad$ $\Sigma f_i \cdot x_i = 29,082.5$

$$\therefore \text{Estimated Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$= \frac{29082.5}{400}$$

$$= \underline{72.7} \text{ minutes}$$

Scanned with CamScanner

§2 Median:

(i) Raw Data: All the terms of the data of n values are arraged in ascending (or descending) order. Then the value of middle most term is defined as median, denoted by M.

(i) If n is odd, then the value of $\left(\frac{n+1}{2}\right)_{th}$ term is median.

(ii) If n is even, then there are two middle values, $\frac{n}{2}_{th}$ & next

$$\therefore M = \frac{\left(\frac{n}{2}_{th} + \text{next value}\right)}{2}$$

Example 19(i) when 'n' is odd,

Given the values of a variable are 3, 2, 5, 4, 7, 5, 6

Let us arrange in ascending order: 2, 3, 4, 5, 5, 6, 7,

here $n = 7$ (odd no).

$$\therefore \text{Med} = \frac{n+1}{2}_{th} = \frac{7+1}{2}_{th} = 4_{th} \text{ value} = \underline{5}$$

(ii) When 'n' is even, 10, 9, 7, 8, 5, 6, 5, 11

arranging in ascending order: 5, 5, 6, 7, 8, 9, 10, 11

here $n = 8$ (even) $\therefore \frac{n}{2} = \frac{8}{2} = 4$

$$\therefore \text{Med} = \frac{4_{th}\text{ Value} + 5_{th}\text{ Value}}{2}$$

$$= \frac{7+8}{2}$$

$$\underline{M = 7.5}$$

Median

§2(ii) Median of ungrouped frequency distribution:

Find $\frac{n}{2}$ (or $\frac{n+1}{2}$), in case n is large, we take

$$Med = \frac{n}{2}_{th} \text{ Value}.$$

for that find cumulative frequencies, and find the value of $\frac{n}{2}$th term.

Example 20: Find the cumulative frequency and hence median.
The spelling test marks are shown in the table bel. --- [3]

Marks	0	1	2	3	4	5
Frequency	2	4	5	5	6	8

S-17/43/Q8(b)

Solution:

Marks	Frequency f_i	Cumulative Frequency C.f
0	2	2
1	4	6
2	5	11
3	5	16
4	6	22
5	8	30

Median class

$n = \Sigma f = 30$

$\frac{n}{2} = \frac{30}{2} = 15$

$Med = 15_{th}$ Value

$= 3$ ✓

§2(iii) Median of Grouped frequency distribution:

To find the median of grouped frequency data of n values.

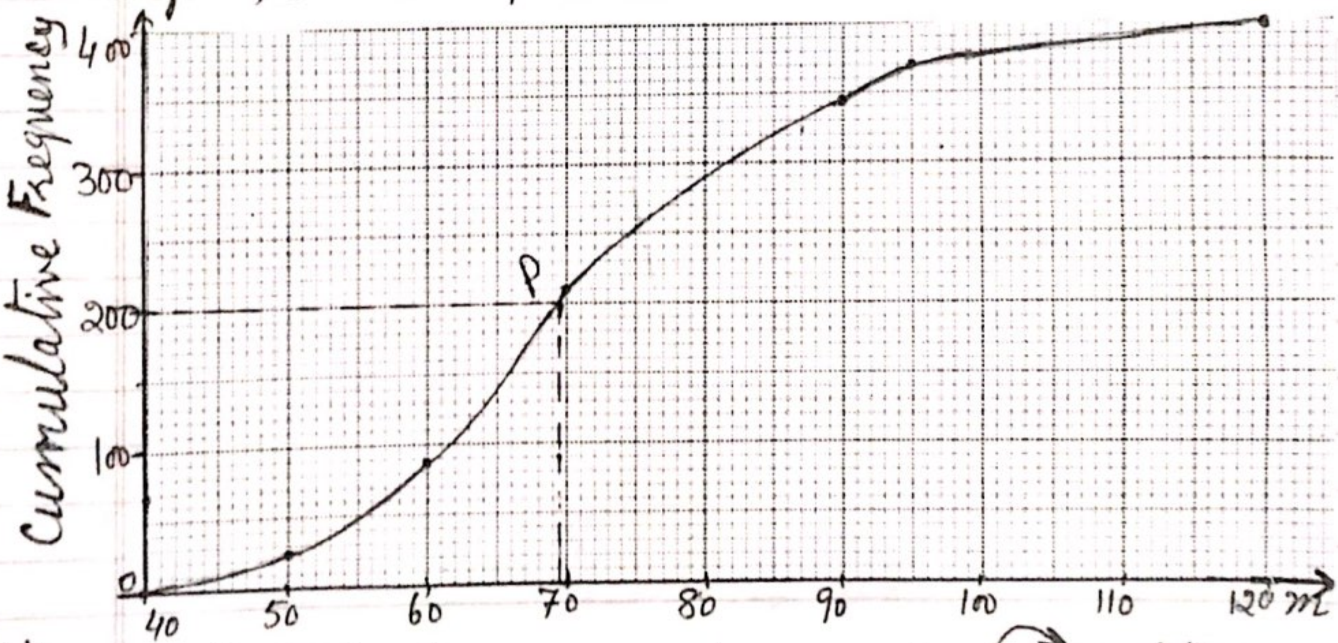
Draw the cumulative frequency curve and then find the value of $\frac{n}{2}$th item.

§3. Cumulative Frequency Graph and Median: Continuous data is represented by a cumulative frequency graph (curve).
Cumulative frequencies are plotted against the upper class boundies of the class intervals.

Example 20: The table shows information about the time taken by 400 people to complete a race.

Time Taken (m - min)	$45 < m \leq 50$	$50 < m \leq 60$	$60 < m \leq 70$	$70 < m \leq 90$	$90 < m \leq 100$	$100 < m \leq 120$
Frequency	23	64	122	136	26	29

(i) On the grid draw the grid draw the cumulative frequence diagram, (ii) and hence find the median. ----[4]



Plot (50, 23), (60, 87), (70, 209), (90, 345), (100, 371), (120, 400)

Time Taken (m - minutes)	Frequency f_i	Time. m	Cumulative frequency C.f
$45 < m \leq 50$	23	$m \leq 50$	23
$50 < m \leq 60$	64	$m \leq 60$	$23+64=87$
$60 < m \leq 70$	122	$m \leq 70$	$87+122=209$
$70 < m \leq 90$	136	$m \leq 90$	$209+136=345$
$90 < m \leq 100$	26	$m \leq 100$	$345+26=371$
$100 < m \leq 120$	29	$m \leq 120$	$371+29=400$

$n = \Sigma f_i = 400$

$$\text{Med} = \frac{n}{2}\text{th} = \frac{400}{2}\text{th} = 200\text{th}$$

$69 < m < 70$ ✓

∴ Median lies between 69 and 70 ✓

Median, Quartiles and Percentiles and Inter-quartile range.

§ 4. Quartiles:

4(i) Lower Quartile 'Q_1' (First Quartile):
If all the terms of the data of n values are arranged in ascending order, then the value of $\frac{n}{4}$th item is defined as lower quartile and is denoted by Q_1.
(or it is the value of item in middle of first half)
(or 25th Percentile $= \frac{25}{100} \times n$th $= \frac{n}{4}$th)

(ii) Upper Quartile 'Q_3' (Third Quartile):

Upper Quartile $= \frac{3}{4}n$th item. (or 75th Percentile)

(or it is the middle item of the second half of the data arranged in ascending order)

(iii) rth Percentile $= \frac{r}{100} \times n$th item.

Example: 30th Percentile $= \frac{30}{100} n$th $= \frac{3n}{10}$th item.

(iv) Inter Quartile Range: upper quartile − Lower quartile

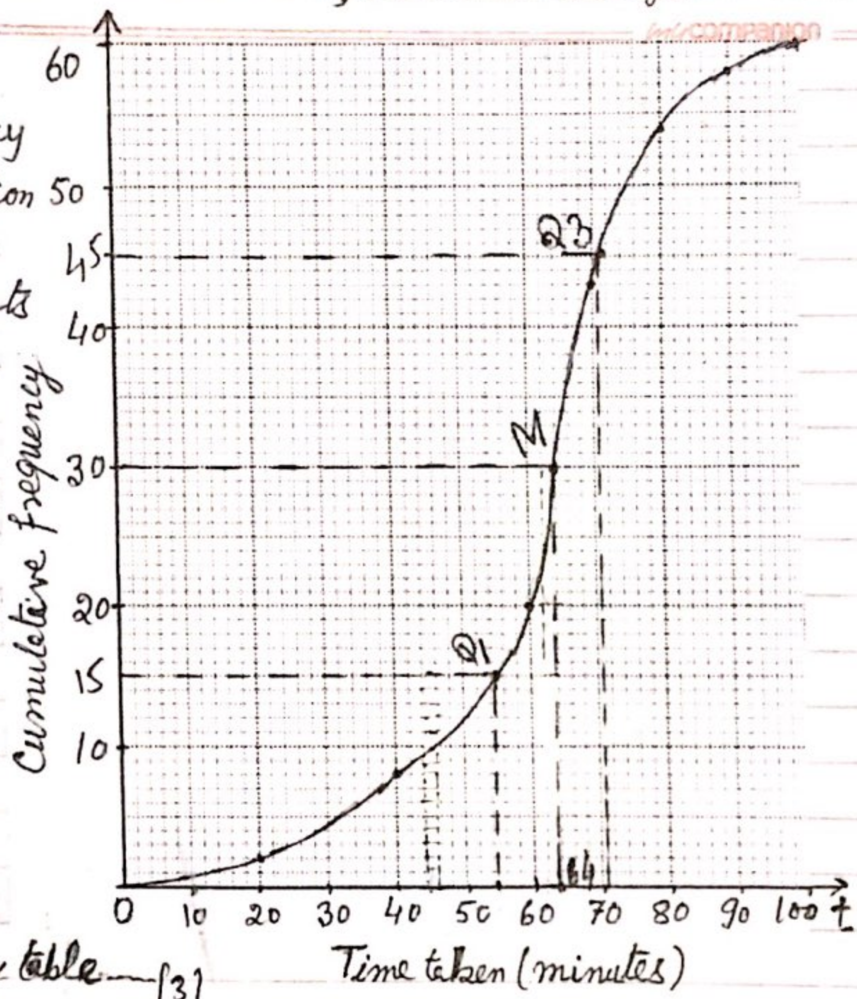
$$= Q_3 - Q_1$$

---

§ Mode: It is the value of an item which repeat maximum number of times.
or
for Grouped data the class-interval which has maximum frequency is called model class.

Example 21:

The cumulative frequency diagram shows information about the time taken, t minutes, by 60 students to complete a test.



(a) Find:

(i) the median. ---[1]

(ii) the inter quartile range. ---[2]

(iii) 40_{th} percentile --[2]

(iv) the number of students who took more than 80 minutes to complete the test. ---[2]

(b) Use the cumulative frequency diagram to complete the frequency table ---[3]

Time taken (t minutes)	$0<t\leq 40$	$40<t\leq 60$	$60<t\leq 70$	$70<t\leq 80$	$80<t\leq 90$	$90<t\leq 100$
Frequency	8				4	

Solution: (a) (i) Med. $(n=60)$ $Med = \frac{60}{2} = 30_{th}$ item value $= 64$ ✓

(ii) $Q_1 = \frac{n}{4}{}_{th} = \frac{60}{4}{}_{th} = 15_{th}$ Value $= 55$

M-16/42/Q4

$Q_3 = 3\frac{n}{4}{}_{th} = \frac{3}{4}\times 60 = 45_{th}$ Value $= 71$

$\therefore$ Inter Quartile range $= Q_3 - Q_1 = 71 - 55 = 16$ ✓

(iii) 40_{th} percentile $= \frac{40}{100}\times 30 = 12_{th}$ item $= 62$ ✓

(iv) Number of students who took more than 80 minutes to comple the test $= (100-54) = 6$ ✓

(b)

$40<t\leq 60$	$60<t\leq 70$	$70<t\leq 80$	$80<t\leq 90$	$90<t\leq 100$
$(20-8)=12$ ✓	$(43-20)=23$ ✓	$(54-43)=11$	4 (given)	$(60-58)=2$ ✓

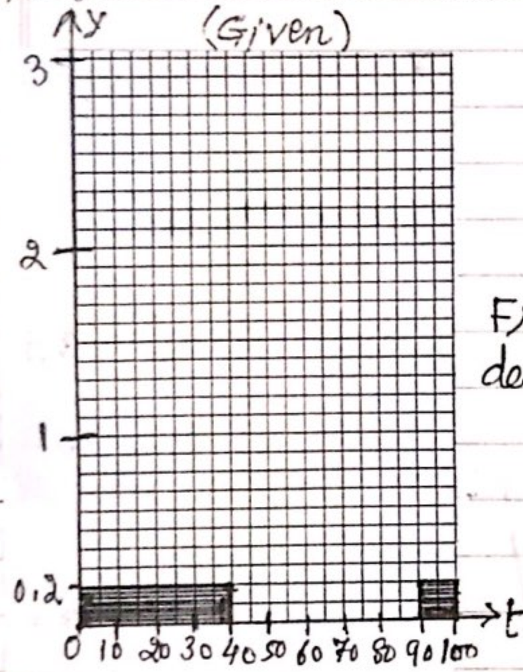
[(c) part continued →]

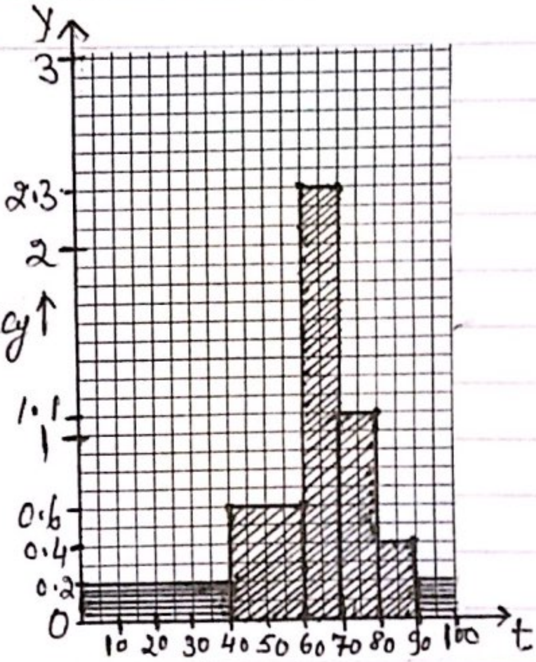
Histogram.
(Continued Example 21)

Example 21(c) On the grid, complete histogram to show the information in table in part(b) ---[4]

Time-t-minute	$0<t\le 40$	$40<t\le 60$	$60<t\le 70$	$70<t\le 80$	$80<t\le 90$	$90<t\le 100$
Frequency	8	12	23	11	4	2

(Given)





Time Taken t-minutes	Frequency f_i	Frequency density f.d
$0<t\le 40$	8	$\frac{8}{40}=0.2$
$40<t\le 60$	12	$\frac{12}{20}=0.6$
$60<t\le 70$	23	$\frac{23}{10}=2.3$
$70<t\le 80$	11	$\frac{11}{10}=1.1$
$80<t\le 90$	4	$\frac{4}{10}=0.4$
$90<t\le 100$	2	$\frac{2}{10}=0.2$

Frequency density:
$$= \frac{\text{Frequency}}{\text{Class size}}$$

Frequency density of the first class interval $=0.2$ is represented by Height 0.2 on the Y-axis

$\therefore$ Scale on Y-axis: 1 Unit = 1 f.d (unit)

M-16/42/Q4

IG-Maths 0580

Statistics P-23

Scatter Diagram and Correlation / Line of Best fit

§5 (i) Scatter Diagram:

It is a plot of points (x, y) in a plane, where x and y are two variable of a given set of Data.

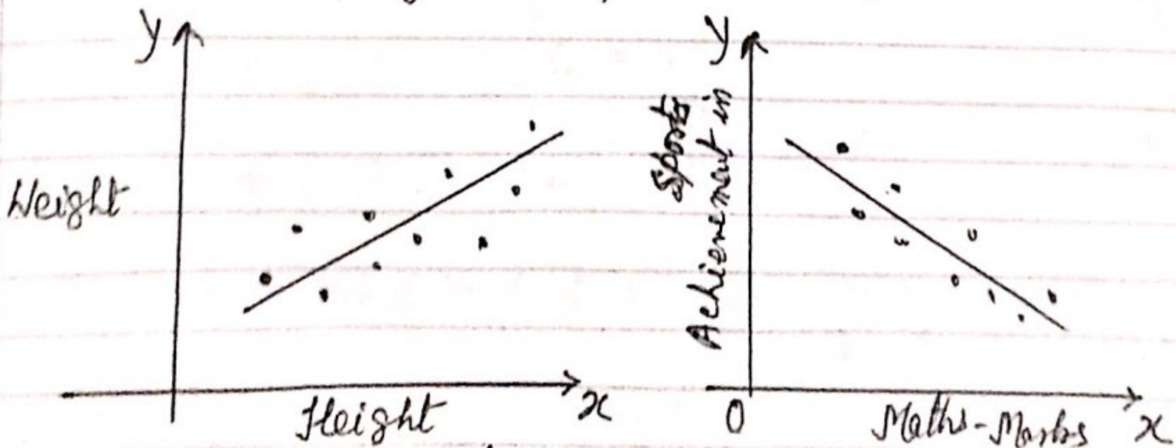
For example, given students of a class, and x denote the height and y denote the corresponding weight of the same student.

§5 (ii) Correlation:

Correlation between two qualities for the same set of data is (i) positive if increase in one than the second also increases in general i.e. the scatter diagram has a upward trend, from left to right, and nega if it has a downward trend.

§5 (iii) Line of Best Fit:

It is solid line that passes between the points on the scatter diagram as close as possible. It passes through $(\bar{x}, \bar{y})$ where $\bar{x}$ and $\bar{y}$ are the means of respectively x-values and y-value for the data.



Positive - Correlation.
(strong)

Scatter Diagram, Correlation, and Line of Best Fit

Example 22: The owners of a small cafe' records the average air temperature and the number of hot drinks he sells each day for a week.

Air temperature (°C)	18	23	19	23	24	25	20
Number of hot drinks sold	12	8	13	10	9	7	12

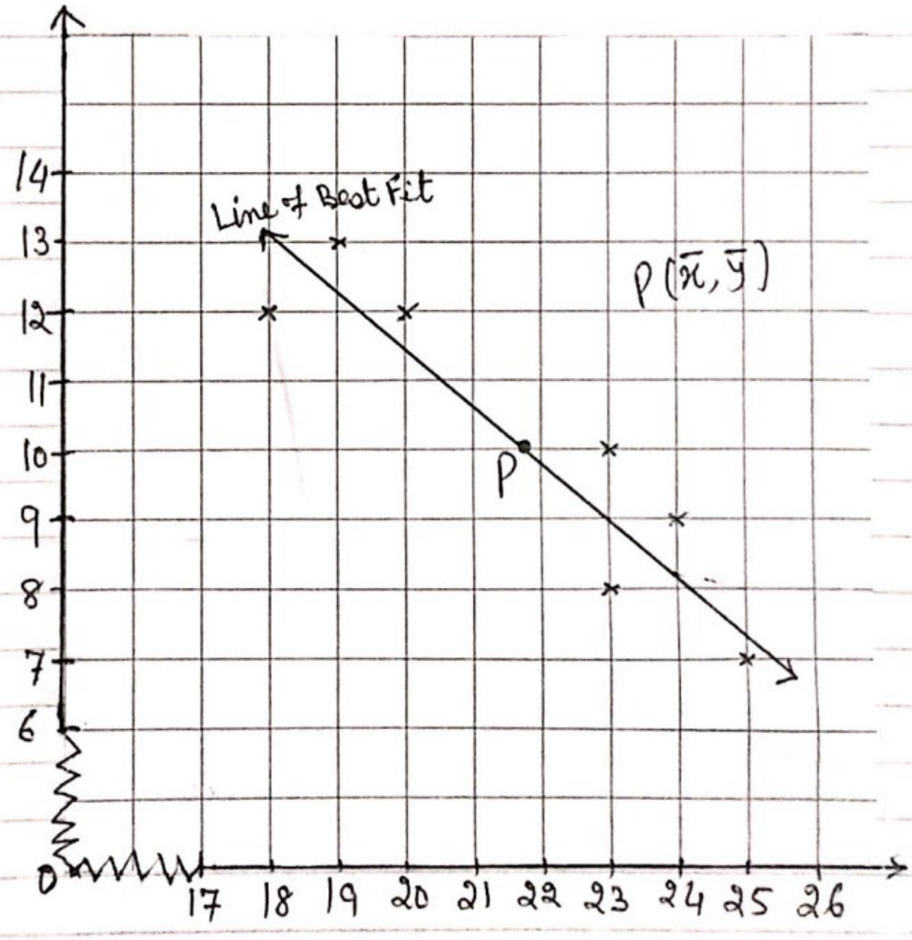
(a) On the grid draw a scatter diagram to show this information. ---[2]

(b) What type of your correlation does your scatter diagram show? --[1]

(c) Draw a line of best fit on the grid. --[1]

S-13/22/Q17

Solution:



(a) 7 correct plots

(b) Negative

(c) ruled line of best fit.

$$\bar{x} = \frac{\Sigma x_i}{7} = \frac{152}{7} = 21.7$$

$$\bar{y} = \frac{\Sigma y_i}{7} = \frac{71}{7} = 10.1$$

Scatter diagram and Correlation

Example 23: The table shows the marks gained by 10 students in their physics test and their mathematics test:

Physics Marks	63	61	14	27	72	75	44	40	28	50
Maths Marks	52	80	16	36	79	75	51	35	24	63

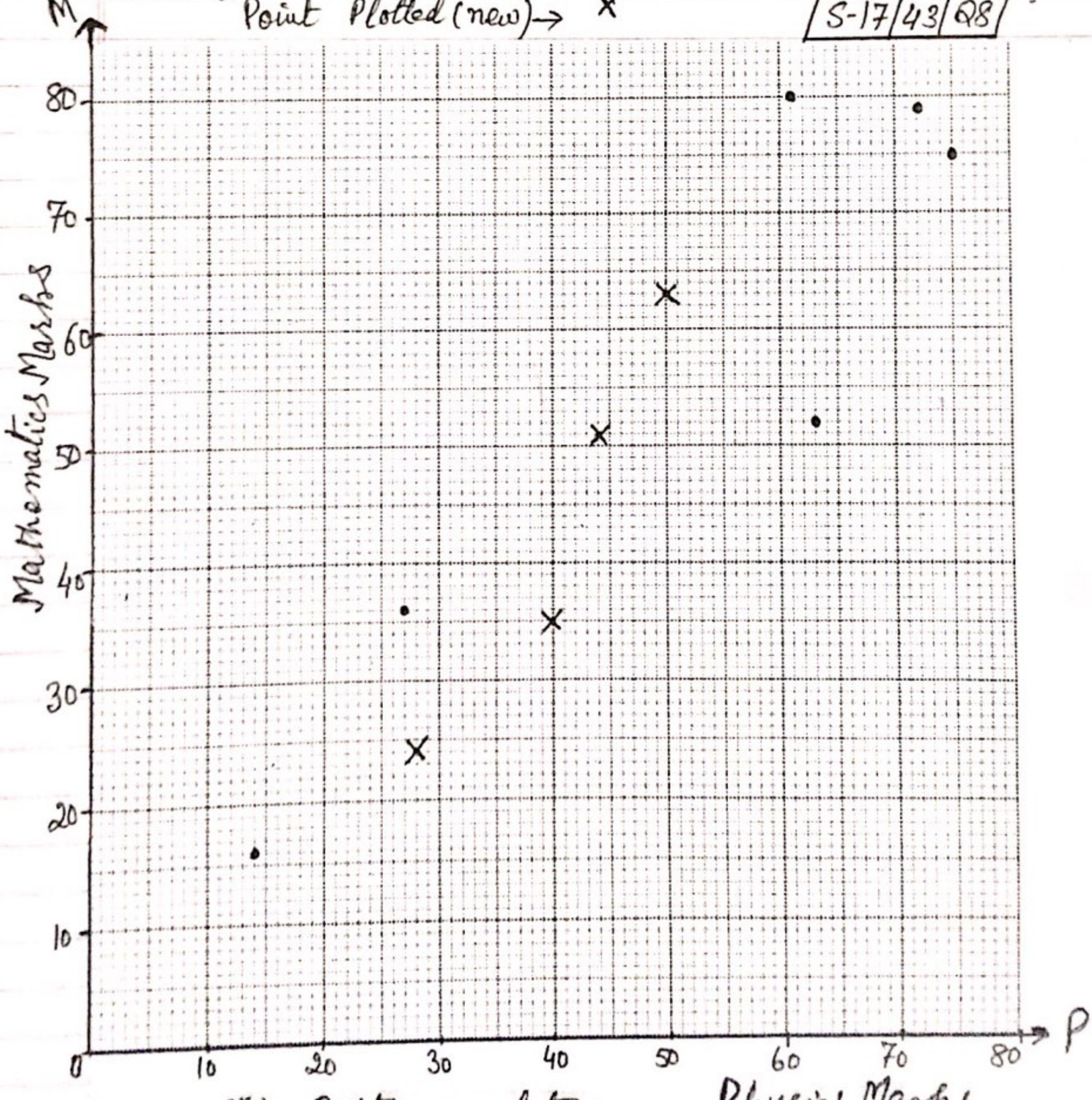
(i) Complete the scatter diagram below. The first six points have been plotted for you. ---[2]

(ii) What type of correlation is shown in the scatter diagram. --[1]

Solution (i) Given Points → •

Point Plotted (new) → ×

S-17/43/Q8



(ii) Positive correlation.