S_1

Statistics and Probability_1

Content. | Page.

1. Stem-and-leaf diagram, Box-and-whisker plot. 1-7
2. Median and Inter quartile range. 8-12
3. Mean, coded-data. 13-18
4. Variance; Combined mean and variance. 19-26
5. Histogram 27-31.

Representation of Data Notes

Suresh Goel
(Former Director)
Alliance World School,
Noida, Delh. NCR.
INDIA.

(+91-9810444804)

Representation of Data - Notes.

§ **Raw Data:** When data is expressed as it was collected, such data is called raw data.

§ **Stem-and-leaf diagram:**

Example 1: The following are the annual amounts of money spent on clothes, to the nearest \$, by 27 people.

10, 40, 60, 80, 100, 130, 140, 140, 140
150, 150, 150, 160, 160, 160, 160, 170, 180
180, 200, 210, 250, 270, 280, 310, 450, 570

Construct a stem-and-leaf diagram for the data above:

Stem	Leaf
0	1 4 6 8
1	0 3 4 4 4 5 5 5 6 6 6 6 7 8 8
2	0 1 5 7 8
3	1
4	5
5	7

Key: 1/4 represents \$ 140
and 0/6 represents \$ 60

Example 2: A random sample of 25 people recorded by number of glasses of water they drink in a particular week. The results are shown below:

23, 19, 32, 14, 25, 22, 26, 36, 45, 42, 47, 28, 17
38, 15, 46, 18, 26, 22, 41, 19, 21, 28, 24, 30.

Draw a stem-and-leaf diagram to represent the data: [W-14/63/Q8]...[3]

Stem	Leaf
1	4 5 7 8 9 9
2	1 2 2 3 4 5 6 6 8 8
3	0 2 6 8
4	1 2 5 6 7

Key 1/4 represents 14 glasses of water.

Example 3: The following are the maximum daily wind speeds in kilometres per hour for the first two weeks in April for two towns, Bronlea and Rogate:

Bronlea	21	45	6	33	27	3	32	14	28	24	13	17	25	22
Rogate	7	5	4	15	23	7	11	13	26	18	26	16	10	34

..[5]

Draw a back-to-back stem-and-leaf diagram: S-16 | 62 | Q5

Bronlea		Rogate
6 3	0	4 5 7 7
5 4 3	1	0 1 3 5 6 8
8 7 5 4 2 1	2	3 3 6
3 2	3	4
5	4	

Key: 3|1|5 represents 13 kph for Bronlea, and 15 kph for Rogate

§ Median of Raw date:

When all the elements of data arranged in ascending (or descending) order, then median is the value of middle most element.

Case I: When 'n' is odd.

The values of a variable are, 3, 2, 5, 4, 7, 4, 6, ; $n=7$

arranging in ascending Order: 2, 3, 4, 4, 5, 6, 7. ;

$$Med = \frac{n+1}{2}th = \frac{7+1}{2}th = 4th \text{ item} = 4$$

Case II:

when 'n' is even: 10, 9, 7, 8, 5, 6, 5, 11 ; $n=8$

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

$$Med = \frac{n+1}{2}th = \frac{8+1}{2}th = 4.5th = \frac{4th + 5th}{2} = \frac{7+8}{2} = 7.5$$

$$Med = 7.5$$

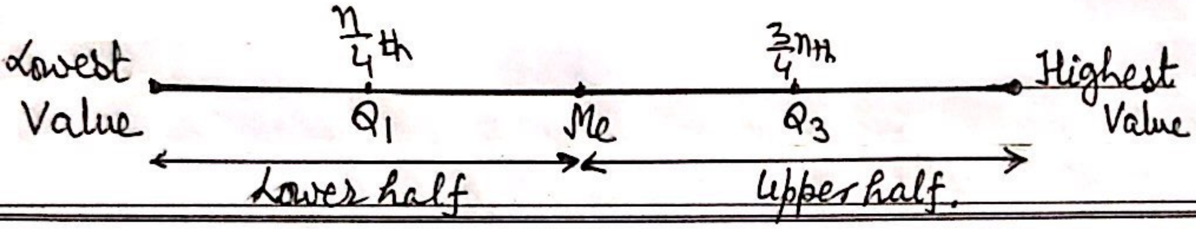
§ Interquartile Range (IQR):

Inter-quartile range = Upper quartile − Lower quartile
$= Q_3 - Q_1$

Lower quartile $Q_1 = \frac{n}{4}$th item

Upper quartile $Q_3 = \frac{3n}{4}$th item.

Median $Me = Q_2$ is the middle most item.



Example 4: Find the quartiles Q_1 and Q_3 and hence find the inter-quartile range (IQR).

(a) 7, 8, 9, 10, 12, 13. ; $n = 6$ (even), divide it into two halves:
Lower half : 7, 8, 9 and upper half. 10, 12, 13.

Q_1 = median of lower half	Q_3 = median of upper half
$Q_1 = 8$	$Q_3 = 12$ ✓

∴ Interquartile range $= Q_3 - Q_1 = 12 - 8 = 4$ ✓

(b) data arranged in order: 5, 6, 7, 8, 9, 10, 10, 18, 19
$n = 9$ (odd) ; $Me = \frac{9+1}{2} = 5_{th} = 9$
$\boxed{\text{Delete } 9}$

Now lower half: 5, 6, 7, 8	Upper half: 10, 10, 18, 19
Q_1 = Med. of lower half $= \frac{1}{2}(6+7) = 6.5$ ✓	Q_3 = Med of upper half $= \frac{10+18}{2} = 14$ ✓

Hence the inter quartile range $= Q_3 - Q_1 = 14 - 6.5$
$= 7.5$ ✓

§ Box-and-whisker plot (Five number summary):

Five-number summary consists of minimum value, the lower quartile Q_1, median Q_2, upper quartile Q_3 and the maximum value.

§ Outliers: Data $> Q_3 + 1.5(Q_3 - Q_1)$ and Data $< Q_1 - 1.5(Q_3 - Q_1)$ are called outliers.

Example 5: The weights in kg, of 15 rugby players in the Rebels club and 15 players in the Sharks club are as shown below:

Rebels	75	78	79	80	82	82	83	84	85	86	89	93	95	99	102
Sharks	66	68	71	72	74	75	75	76	78	83	83	84	85	86	92

(a) Represent the data by drawing a back-back stem-and-leaf diagram with Rebels on the left-hand side of the diagram. --[4]

(b) Find the median and the interquartile range for the Rebels. --[3]

(c) A box-and-whisker plot for the sharks is shown. --[2]
On the same diagram, draw a box-and-whisker plot for the Rebels.

(d) Make one comparison between the weights of the players in the Rebels club and the weights of the players in the Sharks club. --[1]

[W-21/51/Q6]

Solution:

(a)

Rebels		Sharks
	6	6 8
9 8 5	7	1 2 4 5 5 6 8
9 6 5 4 2 2 0	8	3 3 4 5 6
9 5 3	9	2
2	10	

Key: 8|7|2 means 78 kg for Rebels and 72 kg for Sharks.

For Rebels:

(b) Median $= \frac{15+1}{2}$th $= 8$th item $= 84$ kg ✓

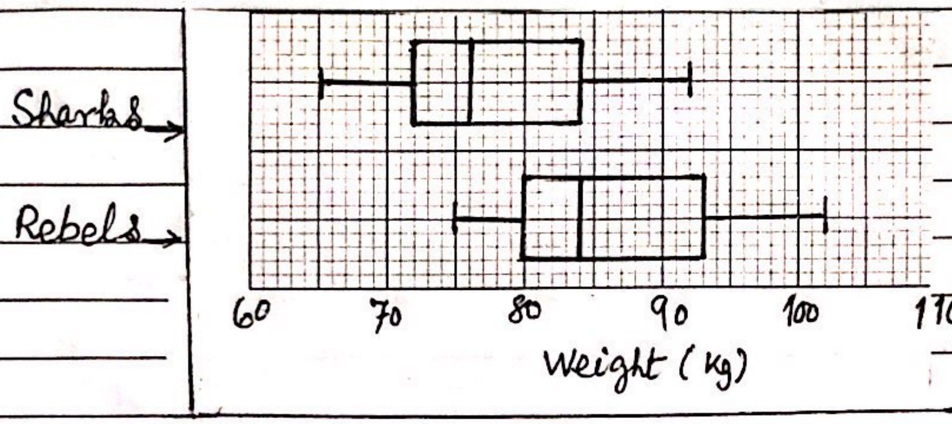
Lower half: 75, 78, 79, 80, 82, 82, 83 $\boxed{84}$ Upper half: 85, 86, 89, 93, 95, 99, 102

$Q_1 = LQ = \frac{7+1}{2}$th $= 4$th item $= 80$ ✓ // $Q_3 = UQ = \frac{7+1}{2}$th $= 4$th item $= 93$ ✓

$IQR = Q_3 - Q_1 = 93 - 80 = \underline{13}$ ✓ (Continued→)

(continued)

5 (c) For Rebels: end points 75 and 102; $Q_1 = 80$, $Me = 84$, $Q_3 = 93$.



(d) Average weight of Rebels is higher than average weight of Sharks.

Example 6: The following back-to-back stem-and-leaf diagram shows the reaction times in seconds in an experiment involving two groups of people, A and B.

	A		B	
(4)	4 2 0 0	20	5 6 7	(3)
(5)	9 8 5 0 0	21	1 2 2 3 7 7	(6)
(8)	9 8 7 5 3 2 2 2	22	1 3 5 6 6 8 9	(7)
(6)	3 7 6 5 2 1	23	4 5 7 8 8 9 9 9	(8)
(3)	8 6 3	24	2 4 5 6 7 8 8	(7)
(1)	0	25	0 2 7 8	(4)

Key 5/22/6 means a reaction time of 0.225 seconds for A and 0.226 seconds for B

(i) Find the median and the interquartile range for group A. ---[3]

The median value for group B is 0.235 seconds, the lower quartile is 0.217 seconds and the upper quartile is 0.245 seconds.

(ii) Draw a box-and-whisker plots for groups A and B on the grid. ---[3]

[W-18/62/ Q2]

For Group A

Solution: (i) $n = 27$, $Me = \frac{27+1}{2}^{th} = 14^{th} = 0.225$ ✓

Lower half - 0.200, --- --- 0.223, $0.\cancel{225}$ → $LQ = \frac{13+1}{2}^{th} = 7^{th} = 0.215$ ✓

Upper half - above $0.\cancel{225}$ → 0.22 --- 0.250 → $UQ = \frac{13+1}{2}^{th} = 7^{th}$ from upper $= 0.236$ ✓

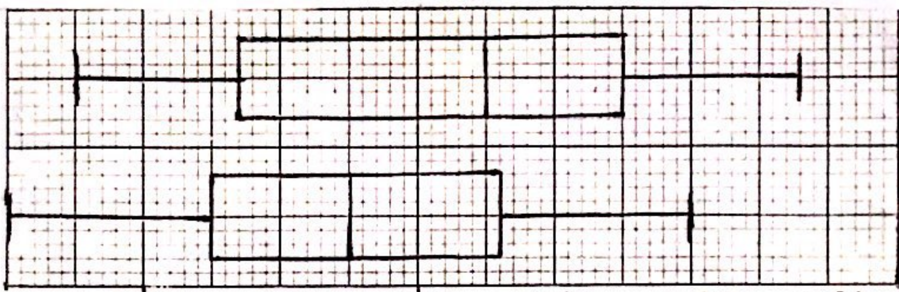
for Group A: $IQR = Q_3 - Q_1 = 0.236 - 0.215 = \underline{0.021}$

(continued→)

Continued:
6(ii)

Group B →

Group A →



	Lowest Value	Q_1	Med	Q_3	Highest Value
A	0·200	0·215	0·225	0·236	0·250
B	0·205	0·217	0·235	0·245	0·258

Example 7: The weights, in kg, of the 11 members of the Dolphin's swimming team and the 11 members of the Sharks swimming team are:

Dolphins	62	75	69	82	63	80	65	65	73	82	72
Sharks	68	84	59	70	71	64	77	80	66	74	72

(i) Draw a back-to-back stem-and-leaf diagram to represent this information, with Dolphins on the left-hand side of the diagram and Sharks on the right-hand side. --[4]

(ii) Find the median and interquartile range for the Dolphins. --[3]

[M-19/62/Q5]

Solution:
(i)

Dolphins		Sharks
	5	9
9 5 5 3 2	6	4 6 8
5 3 2	7	0 1 2 4 7
2 2 0	8	0 4

Key 3|6|4 means 63 kg for Dolphins and 64 kg for Sharks.

For Dolphins:

(ii) $n = 11$, $Me = \frac{11+1}{2}th = 6th$ item $= \underline{72}$ ✓

Lower half: "62, 63, 65, 65, 69" ~~72~~ Upper half "73, 75, 80, 82, 82"

$Q_1 = LQ = \frac{5+1}{2}th = 3rd$ item $= 65$ ✓

$Q_3 = UQ = \frac{5+1}{2}th = 3rd$ item of upper Half $= 80$ ✓

Interquartile range $= Q_3 - Q_1 = 80 - 65 = \underline{15}$ ✓

Example 8: The following are the annual ammounts of money spent on clothes to the nearest of \$10, by 27 people:
10, 40, 60, 80, 100, 130, 140, 140, 140, 150, 150, 150, 160, 160, 160, 160, 170, 180, 180, 200, 210, 250, 270, 280, 310, 450, 570

(i) Construct a stem-and-leaf diagram for the data. ---[3]

(ii) Find the median and the interquartile range of the data. --[3]

An outlier is defined as any data value which is more that 1.5 times the interquartile range above the upper quartile, or more than 1.5 times the interquartile range below the lower quartile. --[3]

[S-13/62/Q5]

Solution:

Stem	Leaf
0	1 4 6 8
1	0 3 4 4 4 5 5 5 6 6 6 6 7 8 8
2	0 1 5 7 8
3	1
4	5
5	7

Key 1/4 represents \$140

(ii) $n = 27$, Median $= \frac{27+1}{2}$ th $= 14_{th}$ item $= \underline{160}$ ✓

Lower half: 10, 40, 60 ----, 150, 160, ~~160~~, $Q_1 = L \cdot Q = \frac{13+1}{2} = 7_{th} = 140$ ✓

Upper half: ~~160~~, 160, 160, 170, ----, 570, $Q_3 = U \cdot Q = \frac{13+1}{2}th = 7_{th} = 210$ ✓

Inter quartile range (IQR) $= Q_3 - Q_1 = 210 - 140 = \underline{70}$ ✓

(iii) To find outliers: $1.5 \times IQR = 1.5 \times 70 = 105$ ✓

Now upper outliers are data $> (105 + 210) \rightarrow$ data > 315
$= \underline{450} \,\&\, \underline{570}$

and the lower outlier $< (140 - 105)$ or < 35
$= \underline{10}$ ✓

Hence the outliers are: $\underline{10, 450, 570}$ ✓

§ Median of Ungrouped frequency distribution:

Find the cumulative frequencies and find the value of $\left(\frac{n+1}{2}\right)_{th}$ or ($\left(\frac{n}{2}\right)_{th}$ item in case n is large)

Example 9:

Variable x	frequency f	c.f
10	4	4
11	7	11
12	8	19
13	11	30
14	10	40
	$n = \Sigma f = 40$	

13 — Median class

$n = 40$

$Med = \frac{40}{2}_{th} = 20_{th}$ item $= \underline{13}$ ✓

$\therefore$ Median $= 13$ ✓

Example 10: Marks in spelling test are shown in the table below: Find the median.

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

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

3 — Median Class.

$n = \Sigma f = 30$

$Med = \frac{30}{2}_{th} = 15_{th}$ Value $= \underline{3}$ ✓

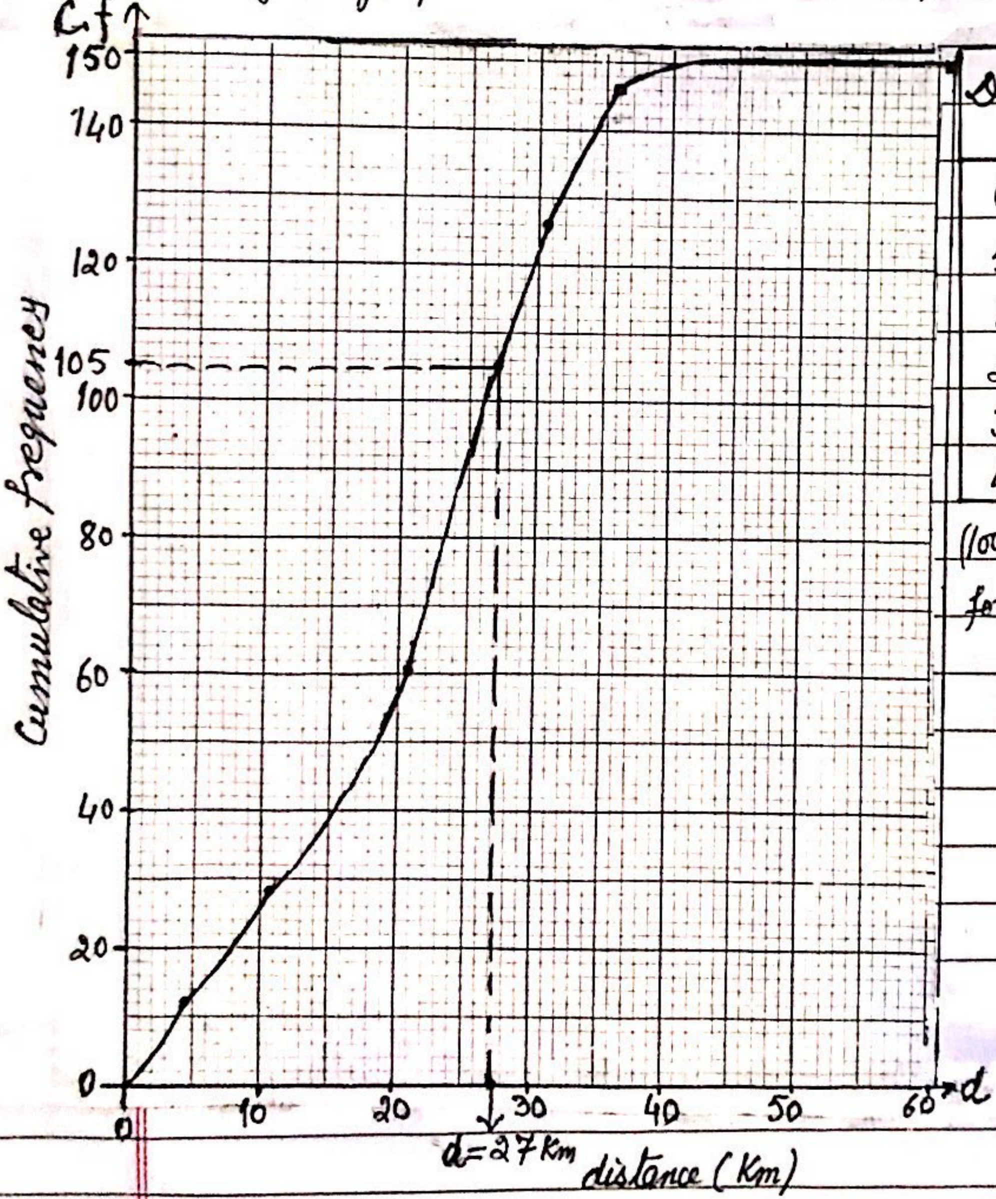
§ Cumulative Frequency Graph and to calculate median and percentile:

Example 11: A driver records the distance travelled in each of 150 journeys. These distances, correct to the nearest km, are summarised in the following table:

Distance (Km)	0-4	5-10	11-20	21-30	31-40	41-60
Frequency	12	16	32	66	20	4

(a) Draw a cummulative frequency graph to illustrate the data. ---[4]

(b) For 30% of these journeys the distance travelled is d Km or more. Use your graph to estimate the value of d. [M-21/52/Q5] ---[2]



Distance	upper boundary	C.f
0-4	4.5	12
5-10	10.5	28
11-20	20.5	60
21-30	30.5	126
31-40	40.5	146
41-60	60.5	150

$(100-30) = 70\%$ of $150 = 105$

for C.f 105 → distance =

d =

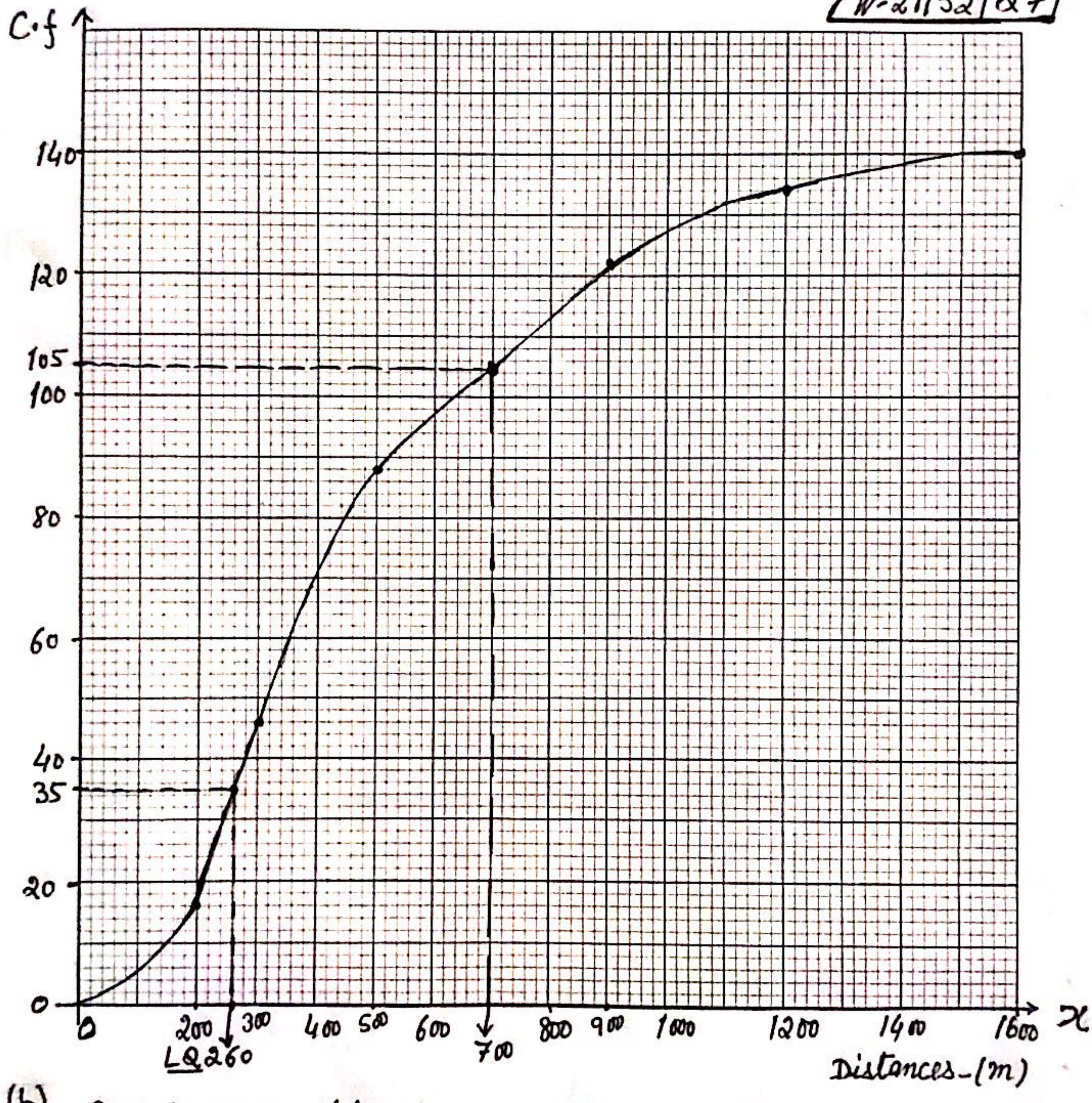
12. The distances, x m, travelled to school by 140 children were recorded. The results are summarised in the table below.

Distance, x m	$x \leqslant 200$	$x \leqslant 300$	$x \leqslant 500$	$x \leqslant 900$	$x \leqslant 1200$	$x \leqslant 1600$
Cumulative frequency	16	46	88	122	134	140

(a) On the grid, draw a cumulative frequency graph to represent these results. [2]

(b) Use your graph to estimate the interquartile range of the distances --[2]

W-21/52/Q7



(b) Q_1 = Lower quartile $= \frac{n}{4}$th $= \frac{140}{4}$th $= 35$th $= 260$

Q_3 = Upper quartile $= \frac{3}{4}n$th $= \frac{3}{4}\times 140$th $= 105$th $= 700$

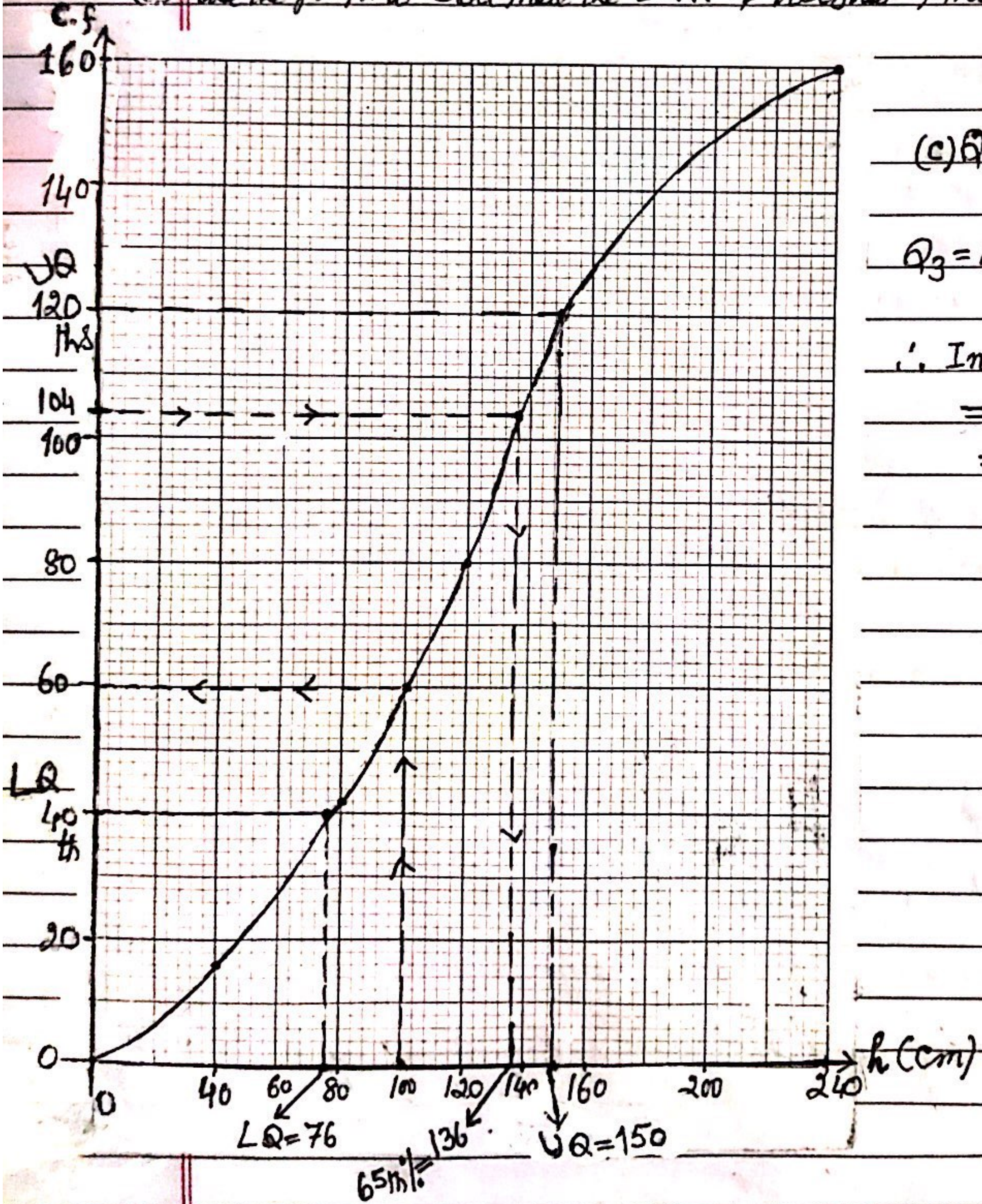
Interquartile range $= Q_3 - Q_1 = 700 - 260 = 440$ ✓

Example 13: The heights in cm of 160 sunflower plants were measured. The results are summarised on the following cumulative frequency curve. [S-21/53/Q1]

(a) Use the graph to estimate the number of plants with heights less than 100 cm. --[1]

(b) Use the graph to estimate the 65_{th} percentile of the distribution. --[2]

(c) Use the graph to estimate the 'IQR' of heights of these plants. --[2]



(c) $Q_1 = LQ = \frac{25}{100} \times 160 = 40_{th}$
$= 76$ ✓

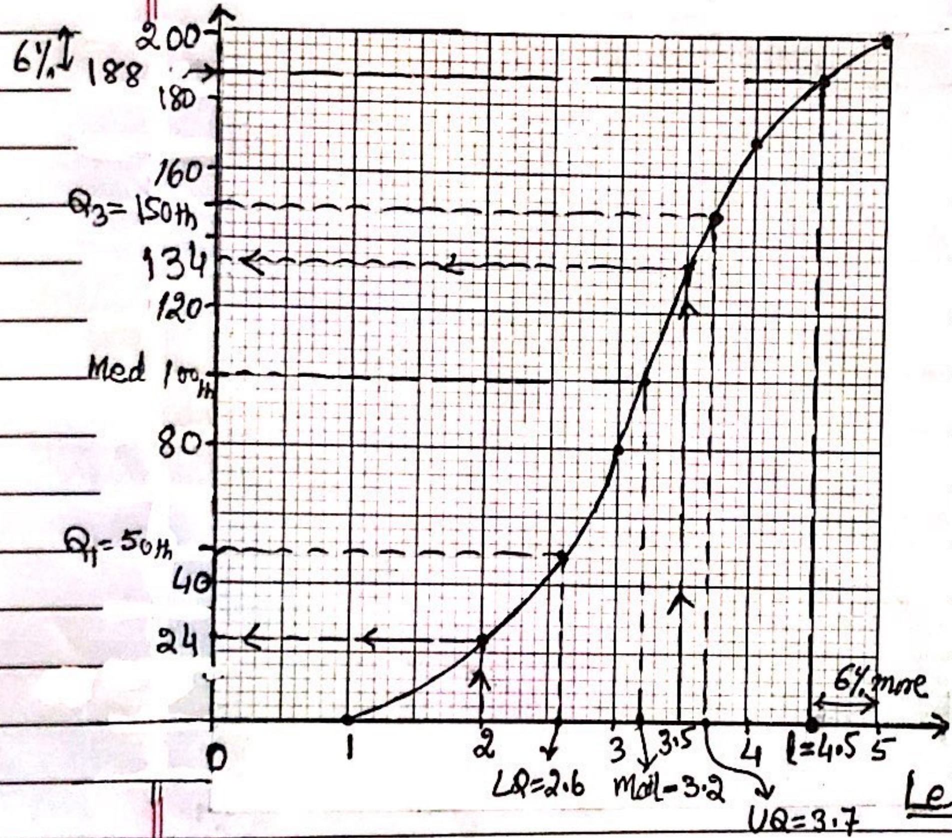
$Q_3 = UQ = \frac{75}{100} \times 160 = 120_{th}$
$= 150$

∴ Interquartile range
$= Q_3 - Q_1$
$= 150 - 76$
$= 74$ cm ✓

(a) Number of plants with heights less than 100 cm $= 60$ ✓

(b) 65_{th} percentile $= 65\%$ of $140 = \frac{65}{100} \times 140 = 104_{th} = 136$ cm.

Example 14: Anabel measured the lengths, in centimetres, of 200 caterpillars. Her results are given in the cumulative frequency graph.
(i) Estimate the median and the interquartile range of the lengths. --[3]
(ii) Estimate how many caterpillars had a length between 2 and 3.5 cm. --[1]
(iii) 6% of caterpillars were of length l cm or more. Estimate l. --[3]
[S-17/62/Q2]



(i) $Med = \frac{1}{2} \times 200_{th} = 100_{th}$ item $= \underline{3.2\text{ cm}}$ ✓
$LQ = Q_1 = \frac{1}{4} \times 200 = 50_{th} = 2.6$ ✓
$UQ = Q_3 = \frac{3}{4} \times 200 = 150_{th} = 3.7$
Interquartile range $= Q_3 - Q_1$
$= 3.7 - 2.6 = \underline{1.1}$ ✓

(ii) No of caterpillars less than 2 = 24
No of caterpillars less than 3.5 = 134
Number of caterpillars between 2 cm and 3.5 cm
$= 134 - 24 = \underline{110}$ ✓

(iii) 6% of 200 = 12 ⇒ 6% more than l = 200 − 12
= 188 to 200
for $188_{th} \rightarrow \underline{l = 4.5}$ ✓

Scanned with CamScanner

§ Mean of Raw-data:

Given n values: $x_1, x_2, x_3, --- x_n$

$$\text{Mean } \bar{x} = \frac{x_1 + x_2 + -- + x_n}{n} = \frac{\Sigma x_i}{n}$$

$$\bar{x} = \frac{\Sigma x_i}{n} \quad [\text{or Sum of data; } \Sigma x_i = n \cdot \bar{x}]$$

Example 15: The age at which a child first walked (to the nearest month) for 8 children. The results were as follows: 12, 11, 16, 19, 10, 12, 12, 13; find the mean age.

Solution: $\text{Mean } \bar{x} = \frac{\Sigma x_i}{n} = \frac{12+11+16+19+10+12+12+13}{8} = \frac{105}{8} = \underline{13 \cdot 125}$

$\therefore \bar{x} = 13$ (To the nearest months)

§ Short cut method of finding mean of raw data (using coded-date)

From each data we subtract a suitable positive number 'a' (assumed mean) and obtain the coded data $(x_i - a)$

Then, $\bar{x} = a + \frac{\Sigma(x_i - a)}{n}$ {Here $\frac{\Sigma(x_i-a)}{n}$ is the Coded mean},

$$\left[\frac{\Sigma(x_i - a)}{n} = \frac{\Sigma x_i - na}{n} = \frac{\Sigma x_i}{n} - a = \bar{x} - a\right.$$

Example 15': x_i: 12, 11, 16, 19, 10, 12, 12, 13

Let assumed mean $a = 10$

$(x_i - 10) =$ 2, 1, 6, 9, 0, 2, 2, 3

$$\Sigma(x_i - 10) = 2+1+6+9+0+2+2+3 = 25$$

$$\therefore \bar{x} = a + \frac{\Sigma(x_i - a)}{n} = 10 + \frac{25}{8} = 10 + 3 \cdot 125 = 13 \cdot 125$$

$\therefore \text{Mean } \bar{x} = \underline{13}$ (To the nearest month)

Example 16: A summary of 24 observations of x is such that: $\Sigma(x-a) = -72.3$ and the mean of these values of x is 8.95. Find the value of the constant a. ---[2]

[W-07/Q1]

Solution: Mean $\bar{x} = a + \frac{\Sigma(x-a)}{n}$ $\left\{ \begin{array}{l} n=24, \quad \bar{x}=8.95 \\ \Sigma(x-a)=-72.3 \end{array} \right.$

$$\Rightarrow \quad 8.95 = a + \frac{(-72.3)}{24}$$

$$\Rightarrow \quad a = 8.95 + \frac{72.3}{24} = 8.95 + 3.0125$$
$$= 11.9625 \checkmark$$

Example 17: Given $\Sigma(t-35) = -15$ for 12 observations, find mean

[S-2007/Q1] --[2]

Solution: Mean $\bar{x} = a + \frac{\Sigma(t-a)}{n}$ $\left\{ \begin{array}{l} \Sigma(t-35) = -15 \\ \Rightarrow a = 35 \\ n = 12 \end{array} \right.$

$$\Rightarrow \bar{x} = 35 + \frac{(-15)}{12}$$

$$\bar{x} = 35 - 1.25 = \underline{33.75} \checkmark$$

Example 18: A summary of 70 observations is given by: $\Sigma(x-60) = 245$

(i) Find the mean. ---[2]

(ii) Find $\Sigma(x-50)$ --[2]

[W-10/63/Q4]

Solution: Given $n = 70$ and $\Sigma(x-60) = 245$

(i) $\bar{x} = a + \frac{\Sigma(x-a)}{n}$ $\} \Rightarrow a = 60$

$$= 60 + \frac{245}{70}$$

$$= 60 + 3.5$$

$$\bar{x} = \underline{63.5} \checkmark \quad ---(1)$$

(ii) Now $\bar{x} = 63.5$, $n = 70$

To find $\Sigma(x-50)$

$$\Sigma(x-50) = \Sigma x_i - n \times 50$$
$$= n.\bar{x} - 50 \times n$$
$$= 70 \times 63.5 - 50 \times 70$$
$$= 4445 - 3500$$
$$= \underline{945} \checkmark$$

§ Mean of Ungrouped frequency distribution:

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{\Sigma f_i} = \frac{\Sigma f_i \cdot x_i}{\Sigma f_i}$$

Example 19: Find the mean;

Marks x_i	f_i	$f_i x_i$
3	4	12
4	5	20
5	2	10
6	1	6
7	3	21
	$\Sigma f = 15$	$\Sigma f_i x_i = 69$

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{69}{15} = 4.6 \checkmark$$

§ Mean of Grouped frequency distribution (Estimated mean):

Example: 20.

Class Intervals	1-9	10-18	19-27	28-36	37-45	46-54
Frequency	2	5	7	1	3	4

Class Interval	Frequency f_i	Mid Value (or Class Marks) x_i	$f_i x_i$
1-9	2	$\frac{1+9}{2} = 5$	10
10-18	5	$\frac{10+18}{2} = 14$	70
19-27	7	23	161
28-36	1	32	32
37-45	3	41	123
46-54	4	50	200
	$n = \Sigma f_i = 22$		$\Sigma f_i \cdot x_i = 596$

$$\text{Mean } \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f} = \frac{596}{22} = 27.09 \checkmark$$

Note: To find the mean of grouped data, we use the mid point of each class-interval and not the actual raw-data. Hence we call it estimated mean.

Scanned with CamScanner

§ Combined Mean:

Given for data x_i: mean $= \bar{x}$ and number $= n_1$
and for data y_i: mean $= \bar{y}$ and number $= n_2$

Then the combined mean $\bar{z} = \dfrac{n_1 \cdot \bar{x} + n_2 \cdot \bar{y}}{(n_1 + n_2)}$

Example 21: A group of 10 married couples and 3 single men found that the women mean age $\bar{x}_W$ of the 10 woment was 41·2 years and for the 13 men, the mean age $\bar{x}_m$ was 46·3 years. Find the mean age of whole group of 23 people. --[2]
[W-2005/Q4]

Solution: No. of women $n_1 = 10$ and their mean $\bar{x}_W = 41.2$ year
No of mean $n_2 = 13$ and their mean $\bar{x}_m = 46.3$ year

$\therefore$ mean age of the whole group $= \dfrac{n_1 \times \bar{x}_W + n_2 \cdot \bar{x}_m}{n_1 + n_2}$

$= \dfrac{13 \times 41.2 + 10 \times 46.3}{13 + 10} = \dfrac{535.6 + 463}{23}$

$\therefore$ Mean of the whole grp. of 23 people $= 43.4$ (3sf)

Example 22: The ages, x years, of 18 people attending an evening class are summarised as: $\Sigma x = 745$.

(i) Calculate the mean. --[1]

(ii) One person leaves the group and the mean age of the remaining 17 people is exactly 41 years. Find the age of the person who left. [W-2004/Q4] --[2]

Solution: $n = 18$, $\Sigma x = 745$

(i) $\therefore$ Mean $\bar{x} = \dfrac{\Sigma x}{18} = \dfrac{745}{18}$

$\bar{x} = 41.39$

(ii) for $n = 18$ and $\sum_{i=1}^{18} x_i = 745$ --- ①

Now for $n = 17$, $\bar{x} = 41$

$\Rightarrow$ for $n = 17$, $\sum_{i=1}^{17} x_i = 41 \times 17 = 697$ --- ②

$\therefore$ age of the person who left:
$= 745 - 697 = 48$

§ Mean of Grouped Data (continuous):

Example 23: The time taken by 200 players to solve a computer puzzle are summarised in the following table:

Time (t seconds)	$0 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 40$	$40 \leq t < 60$	$60 \leq t < 100$
Number of players	16	54	78	32	20

(i) Calculate an estimate of mean time of these 200 players. ---[2]

(ii) Find the greatest possible value of interquartile range of these times. [S-21/51/Q5] ---[2]

Solution:
(i)

Classes	f_i	Mid Value x_i	$f_i \cdot x_i$	C.f
0 - 10	16	5	80	16
10 - 20	54	15	810	70
20 - 40	78	30	2340	148
40 - 60	32	50	1600	180
60 - 100	20	80	1600	200

LQ lies in

UQ lies in

$\Sigma f = 200$ $\quad \Sigma fx = 6430$

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{6430}{200} = \underline{32.15}$$

(ii) $UQ = \frac{3}{4} \times 200 = 150^{th}$ item = lies in class- (40-60)

$LQ = \frac{1}{4} \times 200 = 50^{th}$ item = lies in class- (10 - 20)

Greatest possible Value of interquatile range $= 60 - 10$
$= \underline{50}$

§ Mean of Grouped data (Discrete):

Example 24: The heights to the nearest metres of 134 office buildings in a certain city are summarised in the table below:

Height (m)	21-40	41-45	46-50	51-60	61-80
Frequency	18	15	21	52	28

Calculate the estimates of mean of these heights. --[2]

W-15/63/Q6

Solution

Classes	Frequency f_i	Class Marks x_i	$f_i \cdot x_i$
21-40	18	30.5	549
41-45	15	43	645
46-50	21	48	1008
51-60	52	55.5	2886
61-80	28	70.5	1974

$\Sigma f_i = 134$ $\Sigma f_i x_i = 7062$

Estimated Mean. $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f}$

$= \frac{7062}{134}$

$= \underline{52.7}$ ✓ (3 sf)

§ Measured of Spread:

1 The range is the difference between the lowest value and the hight value.

2. Variance and Standard deviation σ:

Standard deviation $= +\sqrt{\operatorname{Var} x}$

Variance $\sigma^2 = \frac{1}{n}\Sigma(x_i - \bar{x})^2$

Short cut formula: $\boxed{\operatorname{Var} x = \frac{1}{n}\Sigma x_i^2 - (\bar{x})^2}$

§ Variance of frequency distribution:

$$\operatorname{Var} x = \frac{1}{n} \cdot \Sigma f_i (x_i - \bar{x})^2 \qquad n = \Sigma f_i$$

Short cut formula: $\boxed{\operatorname{Var} x = \frac{\Sigma f_i x_i^2}{n} - (\bar{x})^2}$: $\bar{x} = \frac{\Sigma f_i x_i}{n}$

§ Assumed mean method (using coded-data)

From each data we subtract a suitable positive number 'c' (assumed mean) and obtain the coded data $(x_i - c)$

Then mean $\bar{x} = c + \frac{\Sigma(x_i - c)}{n}$

($\frac{\Sigma(x_i - c)}{n}$ is the coded mean)

(i) $\boxed{\text{Variance } \sigma^2 = \frac{\Sigma(x_i - c)^2}{n} - \left(\frac{\Sigma(x_i - c)}{n}\right)^2}$

(ii) For frequency distribution:

Mean $\bar{x} = c + \frac{\Sigma f_i (x_i - c)}{n}$

$$\boxed{\text{Variance } \sigma^2 = \frac{\Sigma f_i (x_i - c)^2}{n} - \left(\frac{\Sigma f_i (x_i - c)}{n}\right)^2}$$

Example 25: Find the mean and variance for the following raw data:
6, 7, 10, 12, 13, 4, 8, 12

Solution:

Mean $\bar{x} = \frac{\Sigma x_i}{n} = \frac{6+7+10+12+13+4+8+12}{8} = \frac{72}{8} = 9$ ✓

Now $\Sigma x^2 = 6^2+7^2+10^2+12^2+13^2+4^2+8^2+12^2 = 722$

$\text{Var } x = \frac{\Sigma x_i^2}{n} - (\bar{x})^2 \quad ; n = 8$

$\sigma^2 = \frac{722}{8} - 9^2 = 90.25 - 81 = 9.25$ ✓

Example 26: Find the variance for the following data.

x_i	3	8	13	18	23
f_i	7	10	15	10	6

Solution:

x_i	f_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
3	7	21	63
8	10	80	640
13	15	195	2535
18	10	180	3240
23	6	138	3174
$n = \Sigma f_i = 48$		$\Sigma f_i x_i = 614$	$\Sigma f_i \times x_i^2 = 9652$

$\bar{x} = \frac{\Sigma f_i x_i}{n}$

$\text{Var } x = \frac{\Sigma f_i \cdot x_i^2}{n} - \left(\frac{\Sigma f_i x_i}{n}\right)^2$

$= \frac{9652}{48} - \left(\frac{614}{48}\right)^2$

$= 201.08 - 163.60$

$= 37.47$ ✓

Example 27: The distances, x m, travelled to school by 140 children were recorded. The results are summarised in the table:

Distance, x m	$x \le 200$	$x \le 300$	$x \le 500$	$x \le 900$	$x \le 1200$	$x \le 1600$
Cumulative frequency	16	46	88	122	134	140

Calculate estimates of mean and standard deviation of the distances. [6]

[W-21/52/Q7(c)]

Solution:

Classes	C.f	f_i	mid value x_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
0-200	16	16	100	16×100	16×100^2
200-300	46	30	250	30×250	30×250^2
300-500	88	42	400	42×400	42×400^2
500-900	122	34	700	34×700	32×700^2
900-1200	134	12	1050	12×1050	12×1050^2
1200-1600	140	6	1400	6×1400	6×1400^2
		$\Sigma f_i = 140$		$\Sigma f_i x_i = 70700$	$\Sigma f_i x_i^2 = 50405000$

$$\text{Var}\,x = \frac{\Sigma f_i x_i^2}{n} - \left(\frac{\Sigma f_i x_i}{n}\right)^2$$

$$= \frac{50405000}{140} - \left(\frac{70700}{140}\right)^2$$

$$= 360035.7 - (505)^2$$

$$= 360035.7 - 255025$$

$$= 105010.7$$

$\therefore$ Standard deviation $= \sqrt{\text{Var}\,x}$

$$= \sqrt{105010.7}$$

$$= \underline{324}$$

Example 28: A summary of 40 values of x gives the following information

$$\Sigma(x-k)=520, \qquad \Sigma(x-k)^2=9640$$

where k is a constant.

(a) Given that the mean of these 40 values of x is 34, find the value of k. --[2]

(b) Find the variance of these 40 values of x. --[2]

[W-21/51/Q2]

Solution: Given $\Sigma(x-k)=520$; $n=40$, mean $\bar{x}=34$

(a) Now $\bar{x} = k+\frac{\Sigma(x-k)}{n} = 34 \Rightarrow k+\frac{\overset{13}{\cancel{520}}}{\cancel{40}} = 34$

$$\Rightarrow k+13=34 \Rightarrow \underline{k=21}$$

(b) $Var = \frac{\Sigma(x-k)^2}{n} - \left(\frac{\Sigma(x-k)}{n}\right)^2 = \frac{9640}{40} - \left(\frac{520}{40}\right)^2$

$$= 241-13^2 = \underline{72}$$

Example 29: For 40 values of the variable x, it is given that, $\Sigma(x-c)^2=3099.2$, where c is a constant. The standard deviation of these values of x is 3.2.

(i) Find the value of $\Sigma(x-c)$ ---[3]

(ii) Given $c=50$, find the mean of these values of x. ---[1]

[M-19/62/Q2]

Solution: Given $\Sigma(x-c)^2=3099.2$, $n=40$, $\sigma=3.2$

(i) Variance $\sigma^2 = \frac{\Sigma(x-c)^2}{n} - \left(\frac{\Sigma(x-c)}{n}\right)^2 = 3.2^2$ (Given)

$$\Rightarrow (3.2)^2 = \frac{3099.2}{40} - \left(\frac{\Sigma(x-c)}{40}\right)^2$$

$$\Rightarrow \left(\frac{\Sigma(x-c)}{40}\right)^2 = \frac{3099.2}{40} - 10.24 = 77.48-10.24 = 67.24$$

$$\Rightarrow \frac{\Sigma(x-c)}{40} = \sqrt{67.24}$$

$$\Rightarrow \Sigma(x-c) = 40 \cdot \sqrt{67.24} = \underline{328}$$

(ii) Mean $= c+\frac{\Sigma(x-c)}{n}$

$$= 50+\frac{328}{40}$$

$$= 50+8.2 = \underline{58.2}$$

Example 30: A summary of n values of x give the following information:
$\Sigma(x-20) = 136$, $\Sigma(x-20)^2 = 2888$
The mean of the n values of x is 24.25
(i) Find the value of n. ---[2]
(ii) Find Σx^2. [M-18/62/Q5] --[4]

Solution: (i) Mean $\bar{x} = C + \frac{\Sigma(x-C)}{n} \Rightarrow 24.25 = 20 + \frac{136}{n}$

$\Rightarrow \frac{136}{n} = (24.25-20) = 4.25 \Rightarrow n = \frac{136}{4.25} = \underline{32}$ ✓

(ii) $Var = \frac{\Sigma(x-20)^2}{n} - \left(\frac{\Sigma(x-C)}{n}\right)^2 = \frac{2888}{32} - \left(\frac{136}{32}\right)^2 = 72.19$ ✓

Also $Var = \frac{\Sigma x^2}{n} - (\bar{x})^2 = 72.19 \Rightarrow \frac{\Sigma x^2}{32} - (24.25)^2 = 72.19$

$\Rightarrow \frac{\Sigma x^2}{32} = 72.19 + (24.25)^2 = 660.2525$

$\Rightarrow \Sigma x^2 = 32 \times 660.2525 = \underline{21128}$ ✓

Example 31: Tim measured the arms lengths, x cm, of 20 people in his class. He found that $\Sigma x = 1218$ and standard deviation of x was 4.2. Calculate $\Sigma(x-45)$ and $\Sigma(x-45)^2$. --[3]
[W-17/63/Q2]

Solution: $n = 20$, $\Sigma x = 1218$, S.D $\sigma = 4.2$
$\Sigma(x-45) = \Sigma x - n \times 45 = 1218 - 20 \times 45 = \underline{318}$ ✓ --(i)

$Var = \frac{\Sigma(x-c)^2}{n} - \left(\frac{\Sigma(x-c)}{n}\right)^2 = \sigma^2$.

$Var = \frac{\Sigma(x-45)^2}{20} - \left(\frac{318}{20}\right)^2 = (4.2)^2$ $\left\{\begin{array}{l}\text{from (i)}\\ \Sigma(x-c) = 318\end{array}\right.$

$\Rightarrow \frac{\Sigma(x-45)^2}{20} = 17.64 + 252.81 = 270.45$

$\Rightarrow \Sigma(x-45)^2 = 20 \times 270.45 = 5409$

$\therefore \Sigma(x-45)^2 = \underline{5409}$ ✓

(Given $\Sigma x = 1923$; $\Sigma x^2 = 337221$)

Example 32: The heights of the 11 members of the Anvils are denoted by x cm. It is given that 3 new members whose heights are 166 cm, 172 cm and 182 cm. now join the Anvils.
Find the standard deviation of the heights of all 14 members of the Anvils. ---[4]

[W-18/63/Q7]

Solution: $\sum_{i=1}^{11} x_i = 1923$ and $\sum_{i=1}^{11} x_i^2 = 337221$.

The three new members; 166, 172, 182

$\Rightarrow \sum_{i=1}^{14} x_i = 1923 + (166+172+182) = 2443$ ✓

and $\sum_{i=1}^{14} x_i^2 = 337221 + (166^2 + 172^2 + 182^2) = 427485$ ✓

Now for all 14 members $\text{Var} = \dfrac{\sum_{i=1}^{14} x_i^2}{14} - \left(\dfrac{\sum_{i=1}^{14} x_i}{14}\right)^2$

$$\text{Var} = \frac{427485}{14} - \left(\frac{2443}{14}\right)^2 = 30534.64 - (174.5)^2 = 84.39$$

$\therefore$ Standard deviation $= \sqrt{84.39} = \underline{9.19}$ ✓

Example 33: The time taken, t hours, to deliver letters on a particular route each day is measured on 250 working days. The mean time taken is 2.8 hours. Given that $\Sigma(t-2.5)^2 = 96.1$, find the standard deviation of the times taken. ---[3]

[W-15/63/Q1]

Solution: $n = 250$, mean time $\bar{x} = a + \dfrac{\Sigma(t-a)}{n} \Rightarrow 2.8 = 2.5 + \dfrac{\Sigma(t-2.5)}{250}$

$$\Rightarrow \frac{\Sigma(t-2.5)}{250} = 0.3 \quad \text{---- (1)}$$

Now $\text{Var}\ t = \dfrac{\Sigma(t-2.5)^2}{250} - \left(\dfrac{\Sigma(t-2.5)}{250}\right)^2$

$$\sigma^2 = \frac{96.1}{250} - (0.3)^2 = 0.2948$$

$\therefore$ Standard deviation $= \sqrt{0.2948} = \underline{0.543}$ ✓

Example 34: The Quivers Archery club has 12 junior members and 20 senior members. For the junior members, the mean age is 15.5 years and the standard deviation of the ages is 1.2 years. The ages of the senior members are summarised by $\Sigma y = 910$ and $\Sigma y^2 = 42850$, where y is the age of a senior members in years.

(i) Find the mean age of all 32 members of the club. --- [2]

(ii) Find the standar deviation of the ages of all 32 members of club.

[W-18/62/Q5] -- [4]

Solution (i) Combined mean $= \dfrac{n_1 \times \bar{x} + n_2 \bar{y}}{n_1 + n_2} = \dfrac{12 \times 15.5 + 910}{12 + 20}$ $\left[\Sigma x = n_1 \bar{x}, \; \Sigma y = n_2 \cdot \bar{x}_2\right.$

$= 34.25$ ✓

$$\text{Var } x = 1.2^2 = \frac{\Sigma x^2}{12} - (\bar{x})^2 \Rightarrow 1.44 = \frac{\Sigma x^2}{12} - (15.5)^2$$

$$\Rightarrow \Sigma x^2 = 12\left[1.44 + (15.5)^2\right] = 12\left[1.44 + 240.25\right) = 2900.28$$

Now for the whole group of 32 members. (Let z)

$$\Sigma z^2 = \Sigma x^2 + \Sigma y^2 = 2900.28 + 42850 = 45750$$ ✓

$$\text{Var } z = \frac{\Sigma z^2}{N} - (\bar{z})^2 = \frac{45750}{32} - (34.25)^2$$

$$= 256.63$$

$\therefore$ Standard deviation $= \sqrt{256.63} = 16.019$

$\therefore$ SD $\sigma = 16$ ✓

Example 35: The table shows the mean and standard deviation of the weights of turkeys and geese.

	Number of Birds	Mean (Kg)	Standard Deviation (Kg)
Turkeys	9	7.1	1.45
Geese	18	5.2	0.96

(i) Find the mean of 27 birds. ---[2]

(ii) The weights of individual turkeys are denoted by x_t kg and the weights of individual geese by x_g kg. By first finding Σx_t^2 and Σx_g^2, Find the standard deviation of the weights of all 27 birds. ---[5]

[S-15/61/Q5]

Solution: New mean $= \dfrac{n_1 \times \bar{x}_t + n_2 \bar{x}_g}{n_1 + n_2} = \dfrac{9 \times 7.1 + 18 \times 5.2}{9+18} = \underline{5.83}$ ✓

$$\text{Var } x_t = \frac{\Sigma x_t^2}{n_1} - (\bar{x}_t)^2 \Rightarrow \frac{\Sigma x_t^2}{9} - (7.1)^2 = (1.45)^2$$

$$\Rightarrow \Sigma x_t^2 = 9 \times [(1.45)^2 + (7.1)^2] = 472.6125 \checkmark$$

and $$\text{Var } x_g = \frac{\Sigma x_g^2}{n_2} - (\bar{x}_g)^2 \Rightarrow \frac{\Sigma x_g^2}{18} - (5.2)^2 = (0.96)^2$$

$$\Rightarrow \Sigma x_g^2 = 18[(0.96)^2 + (5.2)^2] = 503.3088 \checkmark$$

$$\therefore \text{New Var} = \frac{\Sigma x_t^2 + \Sigma x_g^2}{n_1 + n_2} - (\text{New mean})^2$$

$$= \frac{472.6125 + 503.3088}{9+18} - (5.83)^2 = 2.117$$

$$\therefore \text{Standard deviation} = \sqrt{2.117} = \underline{1.46} \checkmark$$

§ To draw a histogram of grouped frequency distribution:

A histogram is used to illustrate continuous data, but it can also used to show discrete by finding the class boundaries.

Class intervals are plotted against frequency densities.

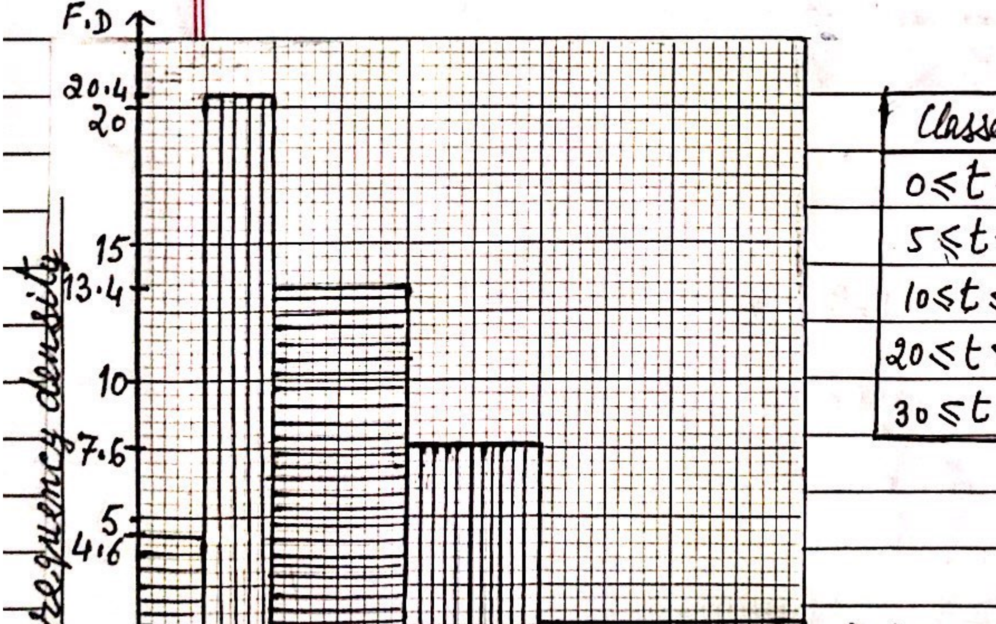
$$\boxed{\text{Frequency density} = \frac{\text{Class frequency}}{\text{Class width}}}$$

Example 36: The times taken, in minutes, by 360 employees at a large company to travel from home to work are shown as:

Time, t minutes	$0 \leq t < 5$	$5 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 30$	$30 \leq t < 50$
Frequency	23	102	135	76	24

[4]

Draw a histogram to represent this information: W-21/53/Q3



Classes	Frequency	Frequency density
$0 \leq t < 5$	23	$\frac{23}{5} = 4.6$
$5 \leq t < 10$	102	$\frac{102}{5} = 20.4$
$10 \leq t \leq 20$	135	$\frac{135}{10} = 13.4$
$20 \leq t < 30$	76	$\frac{76}{10} = 7.6$
$30 \leq t < 50$	24	$\frac{24}{20} = 1.2$

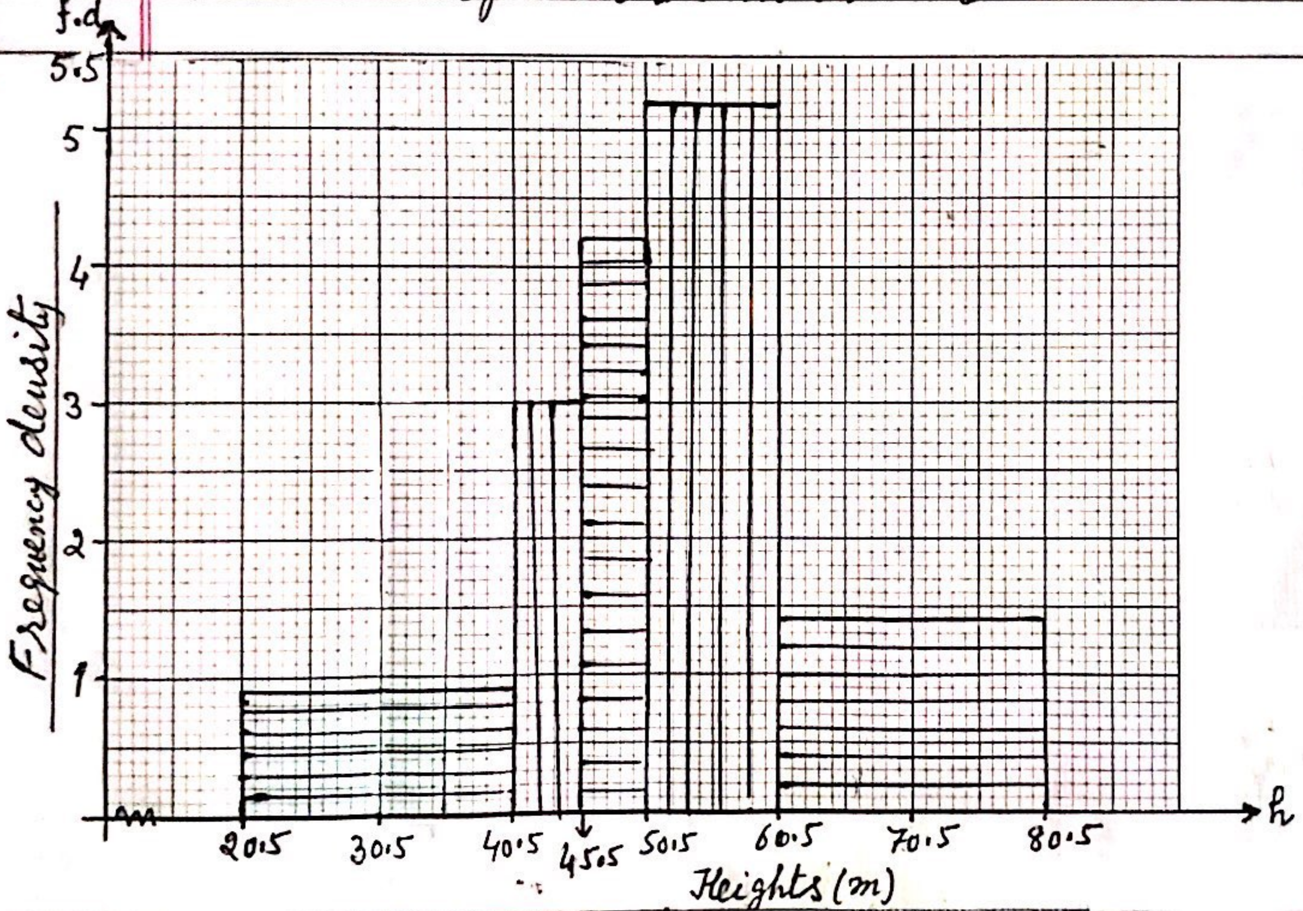
§ Discrete frequency distribution:

Example 37: The heights to the nearest metres of 134 office buildings in a certain city are summarised in the table below:

Heights (m)	21-40	41-45	46-50	51-60	61-80
Frequency	18	15	21	52	28

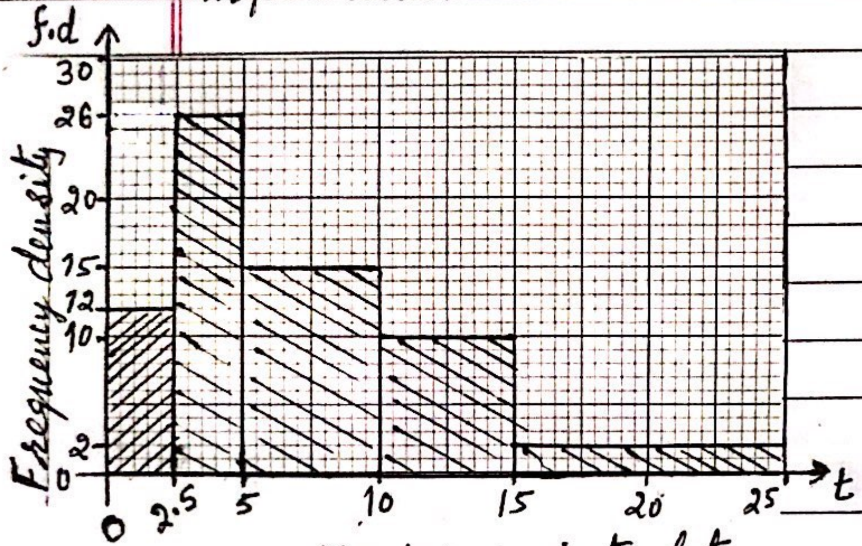
W-15/63/Q6

Draw a histogram to illustrate the data. ---[4]



Class Limits	Class Boundaries	Class Width	Frequency f	Frequency density $= \frac{\text{frequency}}{\text{class width}}$
21-40	20.5-40.5	20	18	$18/20 = 0.9$
41-45	40.5-45.5	5	15	$15/5 = 3$
46-50	45.5-50.5	5	21	$21/5 = 4.2$
51-60	50.5-60.5	10	52	$52/10 = 5.2$
60-80	60.5-80.5	20	28	$\frac{28}{20} = 1.4$

Example 38: 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 trains recorded that arrived more than 10 minutes late. [I.G/S-16/22/Q20] --[2]

Solution: Frequency density $= \dfrac{\text{frequency}}{\text{Class width}} \Rightarrow$ frequency $=$ Frequency density $\times$ class width.

(a) $\therefore$ Total numbers 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 $= 50 + 20 = 70$

Late % $= \dfrac{70}{240} = \underline{29.2}\ \%$

Example 39. The table shows information about the time, t minutes, taken for each of 150 girls to complete an essay.

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

The information in the frequency table is shown in a histogram, the height of the block for $60 < t \leq 65$ interval is 5 cm. Complete the table: ---[3]

Time (t min)	$60 < t \leq 65$	$65 < t \leq 70$	$70 < t \leq 80$	$80 < t \leq 100$	$100 < t \leq 150$
Height of Block cm	5				

[IG-/W-17/43/Q4]

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

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

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

But height $=$ 5 cm.

Scale for the height: 1 Unit of f.d $= \frac{5}{2} = 2.5$ cm

$\therefore$ height $= (f.d \times 2.5)$ cm.

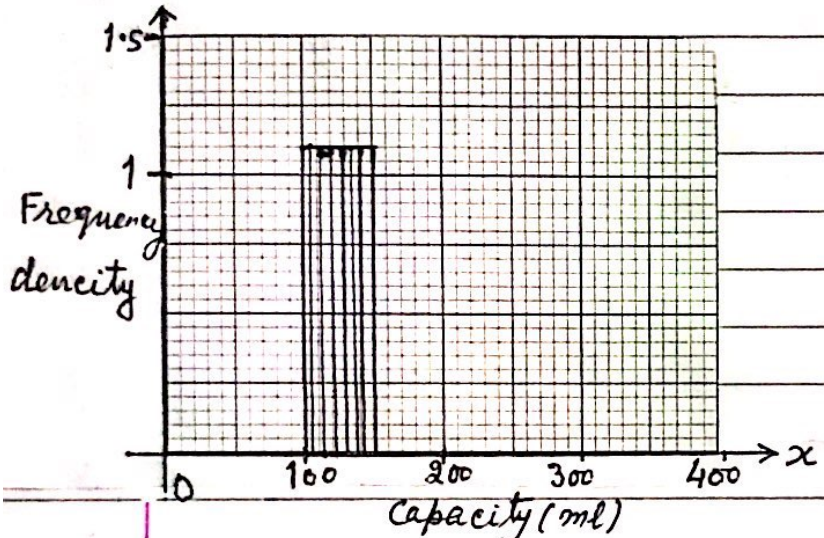
Time (t min.)	$60 < t \leq 65$	$65 < t \leq 70$	$70 < t \leq 80$	$80 < t \leq 100$	$100 < t \leq 150$
Frequency	10	26	34	58	22
Fre. density	$\frac{10}{5} = 2$	$\frac{26}{5} = 5.2$	$\frac{34}{10} = 3.4$	$\frac{58}{20} = 2.9$	$\frac{22}{50} = 0.44$
Height of Block $= f.d \times 2.5$	$2 \times 2.5 = 5$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

Example 40: 200 students estimate the capacity, x millilitres, of a cup. The results are shown in the frequency table:

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

Complete the histogram. ---[4]



Solution:

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
Fre. density	$\frac{20}{100} = 0.2$	$\frac{55}{50} = 1.1$	$66/50 = 1.32$	$\frac{35}{50} = 0.7$	$\frac{24}{150} = 0.16$