

Frequency distribution and graphs.

Types of frequency distribution

- * Ungrouped data
- * Grouped data

Procedure for / steps for constructing grouped data

Step I: Find the highest value and lowest value

Step II: Find the range ($R = \text{Highest} - \text{Lowest}$)

Step III: Select the no. of classes desired

Step IV: Find the width by dividing the no. of classes & rounding up

Step V: Select the lower and upper limit

Step VI: Find the class boundary.

Step VII: Tally the data, find the frequencies and find the cumulative frequencies.

1) In a survey of 20 patients, who smoked, the following data were obtained. Each value represent the no. of cigarettes the patient smoked per day. Construct a frequency distribution using six classes.

10	8	6	14
22	13	17	19
11	9	18	14
13	12	15	15
5	11	16	11

$$\text{Step I: } H = 22, L = 5$$

$$\text{Step II: } R = H - L \\ = (22 - 5) = 17$$

$$\text{Step III: no. of classes} = 6 \text{ (Given)}$$

$$\text{Step IV: } \frac{R}{\text{no. of classes}} = \frac{17}{6} = 2.8 \approx 3$$

Step V: Class interval

(Include) 5 - 8	(exclude) (OR)	5 - 7	(Include)
9 - 12		8 - 10	
13 - 16		11 - 13	
17 - 20		14 - 16	
21 - 24		17 - 19	
25 - 28		20 - 22	

Step VI: 4.5 - 7.5

7.5 - 10.5

Class boundary

10.5 - 13.5

13.5 - 16.5

16.5 - 19.5

19.5 - 22.5

Step VII: Frequencies :-Commulative freq.

2

2

3

5

6

11

5

16

3

19

1

20

Freq. Distribution :-

Class int	Class bound	freq.	Cum. freq.
5 - 7	4.5 - 7.5	2	2
8 - 10	7.5 - 10.5	3	5
11 - 13	10.5 - 13.5	6	11
14 - 16	13.5 - 16.5	5	16
17 - 19	16.5 - 19.5	3	19
20 - 22	19.5 - 22.5	1	20

2 No. of class = 5

12	10	2	15
9	8	1	18
6	8	3	16
3	6	9	15
2	6	9	14
1	2	11	13

Sol: Step I : $H = 18$, $L = 1$

Step II : $R = H - L$

$$18 - 1 = 17$$

Step III : no. of classes = 5

Step IV : $\frac{17}{5} = 3.4 = 3$

Class interval	Class boundary	Freq.	Cumu. Freq.
1 - 4	0.5 - 4.5	7	7
5 - 8	4.5 - 8.5	5	12
9 - 12	8.5 - 12.5	6	18
13 - 16	12.5 - 16.5	5	23
17 - 20	16.5 - 20.5	1	24

3. No. of classes = 5

32 47 51 41 46 30

46 38 34 34 52 48

48 38 43 41 24 24

25 29 33 45 51 32

32 27 23 23 34 35

Sol: I : $H = 52$, $L = 21$

II : $(H-L) = 31$

III : $n = 5$

IV : $\frac{31}{5} = 6.2 = 6$

Class inter.	Class bound	Freq.	Cum. freq.
21 - 27	20.5 - 27.5	6	6
28 - 34	27.5 - 34.5	9	15
35 - 41	34.5 - 41.5	5	20
42 - 48	41.5 - 48.5	7	27
49 - 55	48.5 - 55.5	3	30

4. No. of classes = 8

240 210 220 260 250 195 230 270 325 225
 165 295 205 230 250 210 220 210 230 202
 250 265 230 210 240 245 225 180 175 215
 215 235 245 250 215 210 195 240 240 225
 260 210 190 260 230 190 210 230 185 260

Sol I $H = 325 - 165 = 160$

II $R = 325 - 165 = 160$

III $n = 8$

IV $\frac{160}{8} = 20$

Class int	Class boun	Freq	Cum. freq.
165 - 185	164.5 - 185.5	4	4
186 - 206	185.5 - 206.5	6	10
207 - 227	206.5 - 227.5	14	24
228 - 248	227.5 - 248.5	14	38
249 - 269	248.5 - 269.5	9	47
270 - 290	269.5 - 290.5	1	48
291 - 311	290.5 - 311.5	1	49
312 - 332	311.5 - 332.5	1	50

5. 88 88 110 88 80 69 102 78 70 55
 79 85 80 100 60 90 77 55 75 55
 54 60 75 64 105 56 71 70 65 72

$$\text{Sol I } H = 110, L = 54$$

$$\text{II } R = 110 - 54 = 56$$

$$\text{III } n = 7$$

$$\text{IV } \frac{56}{7} = 8$$

Class int.	Class boun	Freq.	C.freq
54-62	53.5-62.5	7	7
63-71	62.5-71.5	6	13
72-80	71.5-80.5	8	21
81-88	80.5-88.5	4	25
89-96	88.5-96.5	1	26
97-104	96.5-104.5	2	28
105-113	104.5-113.5	2	30

6. 767 770 761 760 771 768 776 771 756 770
 763 760 747 766 754 771 771 778 766 762
 780 750 746 764 769 759 757 753 758 746

$$\text{Sol I } H = 780, L = 746$$

$$\text{II } R = 780 - 746 = 34$$

$$\text{III no. of class} = 5$$

$$\text{IV } \frac{34}{5} = 6.8 = 7$$

Class int	Class boun	Freq.	Cumu freq,
46-52	45.5-52.5	4	4
53-59	52.5-59.5	6	10
60-66	59.5-66.5	8	18
67-73	66.5-73.5	9	27
74-80	73.5-80.5	3	30

6 44 39 37 21 31 170 44 635 30 78

42 6 250 43 10 82 50 181 66 37

$$\text{Sol I } H = 635, L = 6$$

$$\text{II } R = 629$$

$$\text{III } n = 5$$

$$\text{IV } \frac{629}{5} = 125.8 \approx 126$$

Class int	Class bound.	Freq.	Cum-freq.
6 - 132	5.5 - 132.5	16	16
133 - 259	132.5 - 259.5	3	19
260 - 386	259.5 - 386.5	0	19
387 - 513	386.5 - 513.5	0	19
514 - 640	513.5 - 640.5	1	20

- Ungrouped frequency distribution

$$1 \quad X = 1, 2, 1, 3, 3, 4, 1, 2, 3$$

Class val.	Freq.	Cum freq.	Tally	Relative freq.
1	3	3		$3/9 = 0.3$
2	2	5		$2/9 = 0.2$
3	3	8		$3/9 = 0.3$
4	1	9		$1/9 = 0.1$

$$2 \quad X = 5, 7, 8, 5, 7, 9, 8, 6, 5, 7$$

Class value	Freq.	Cum freq.	Tally	Relative freq.
5	3	3		$3/10 = 0.3$
6	1	4		$1/10 = 0.1$
7	3	7		$3/10 = 0.3$
8	2	9		$2/10 = 0.2$
9	1	10		$1/10 = 0.1$

- Graphs.

(i) Histogram :- Histogram is a graph that displays the data by using continuous vertical

Bars of various height to represents the freq. of the classes

Karl Pearson is introduce the histogram in 1891

Procedure

Step I: Draw & label X and Y axis - The Y-axis is always a vertical line & X-axis is always a horizontal line.

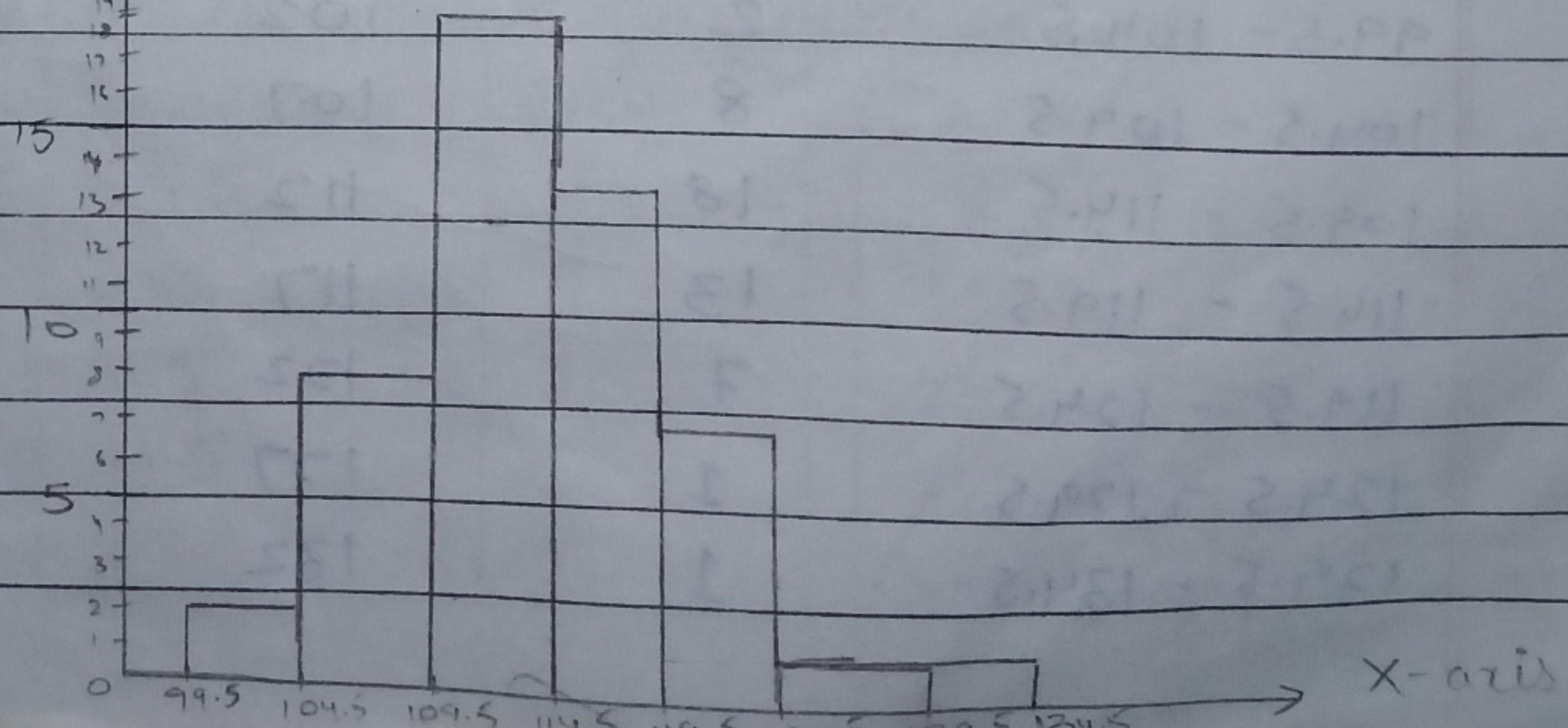
Step II: Represent the frequencies on Y-axis and class boundaries on X-axis.

Step III: Using the frequencies as the heights draw vertical bars for each class.

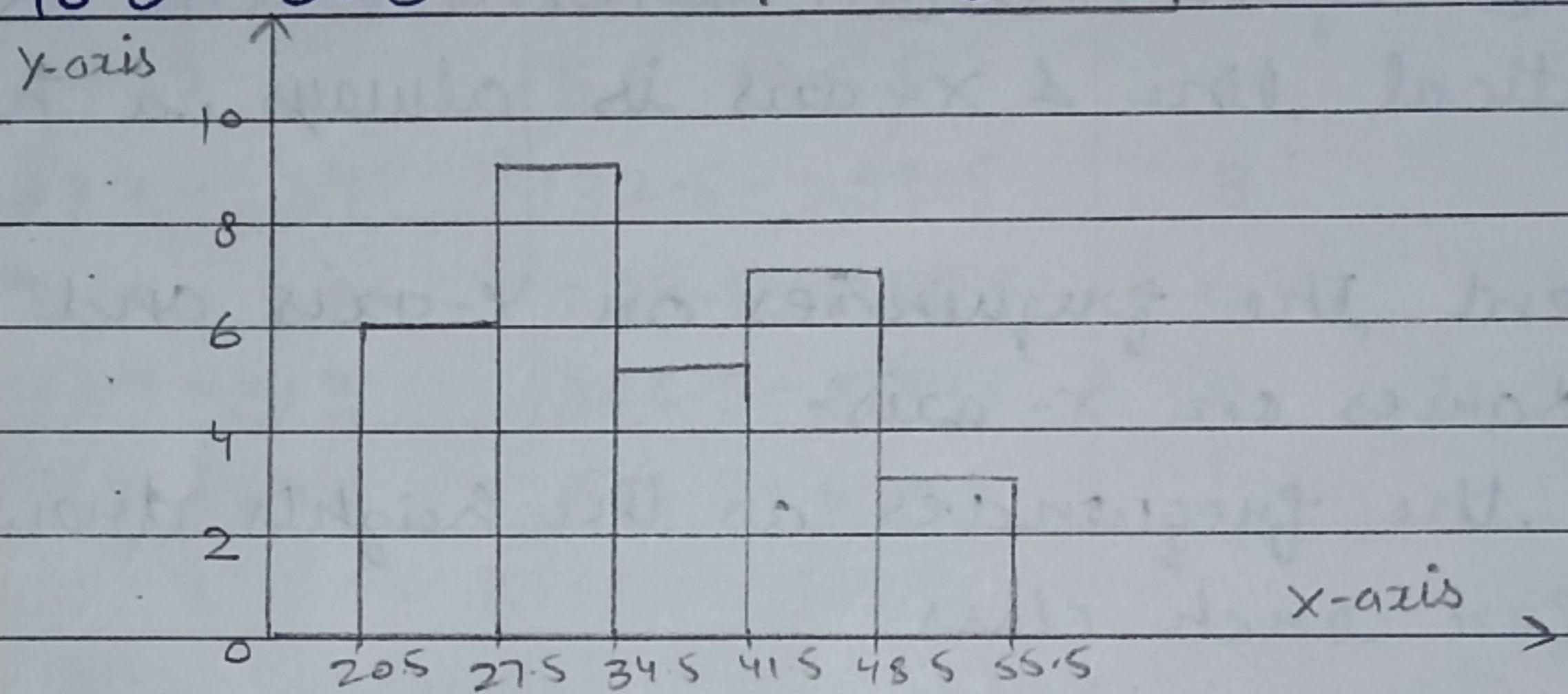
→ Construct a Histogram to represent the data - show for the record of high temperatures for each of 50 - states.

→ Class boundary	Frequency
99.5 - 104.5	2
104.5 - 109.5	8
109.5 - 114.5	18
114.5 - 119.5	13
119.5 - 124.5	7
124.5 - 129.5	1
129.5 - 134.5	1

Y-axis 20



Class boundary	Freq.
20.5 - 27.5	6
27.5 - 34.5	9
34.5 - 41.5	5
41.5 - 48.5	7
48.5 - 55.5	3



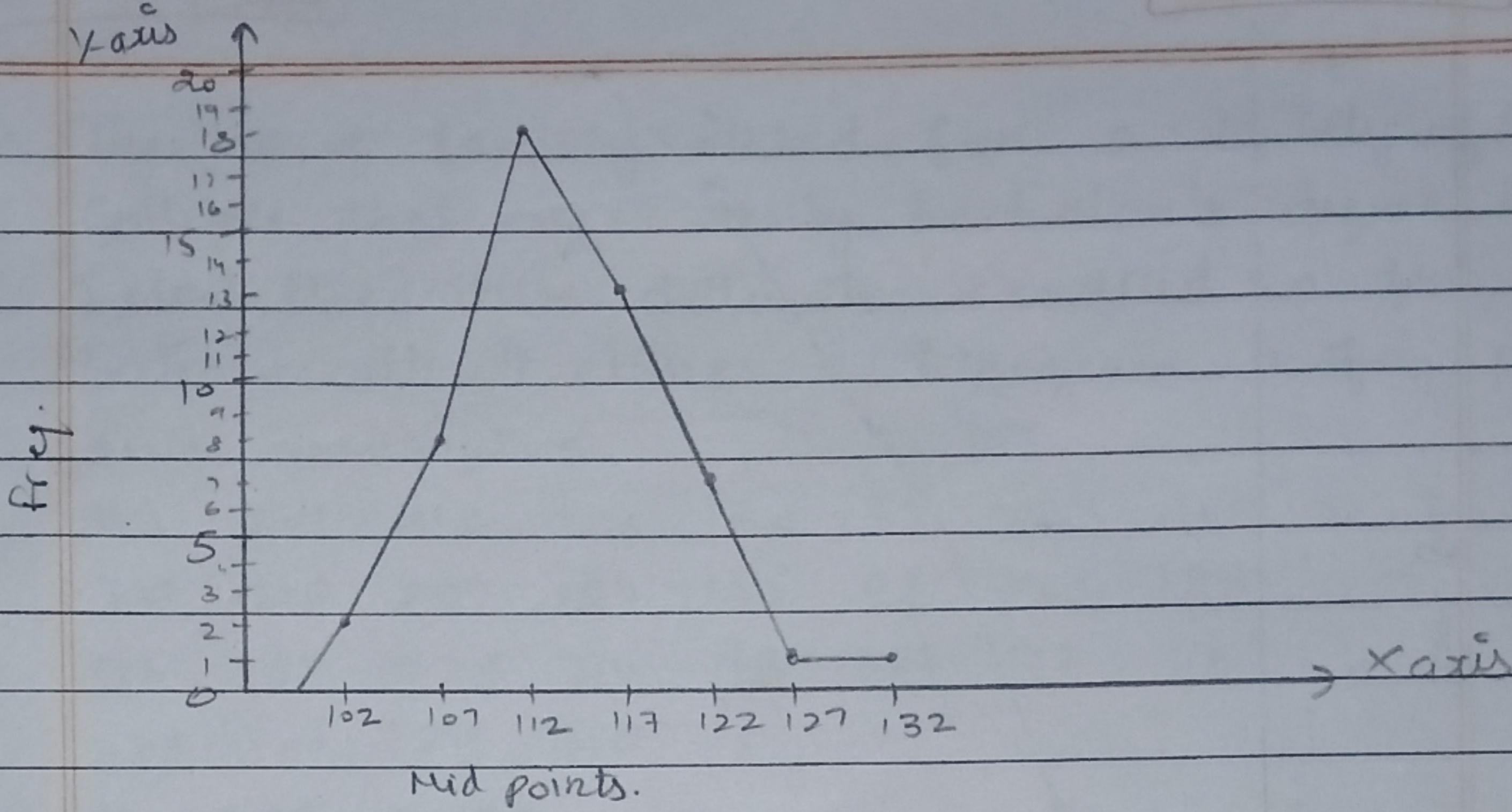
- Frequency Polygon :- It is a graph that displays the data by using lines that connect points plotted for the frequencies at the mid point of the classes. The frequencies are represented by the heights of the points.

Step I : Find the mid point :- upper b.t lower boundary / 2

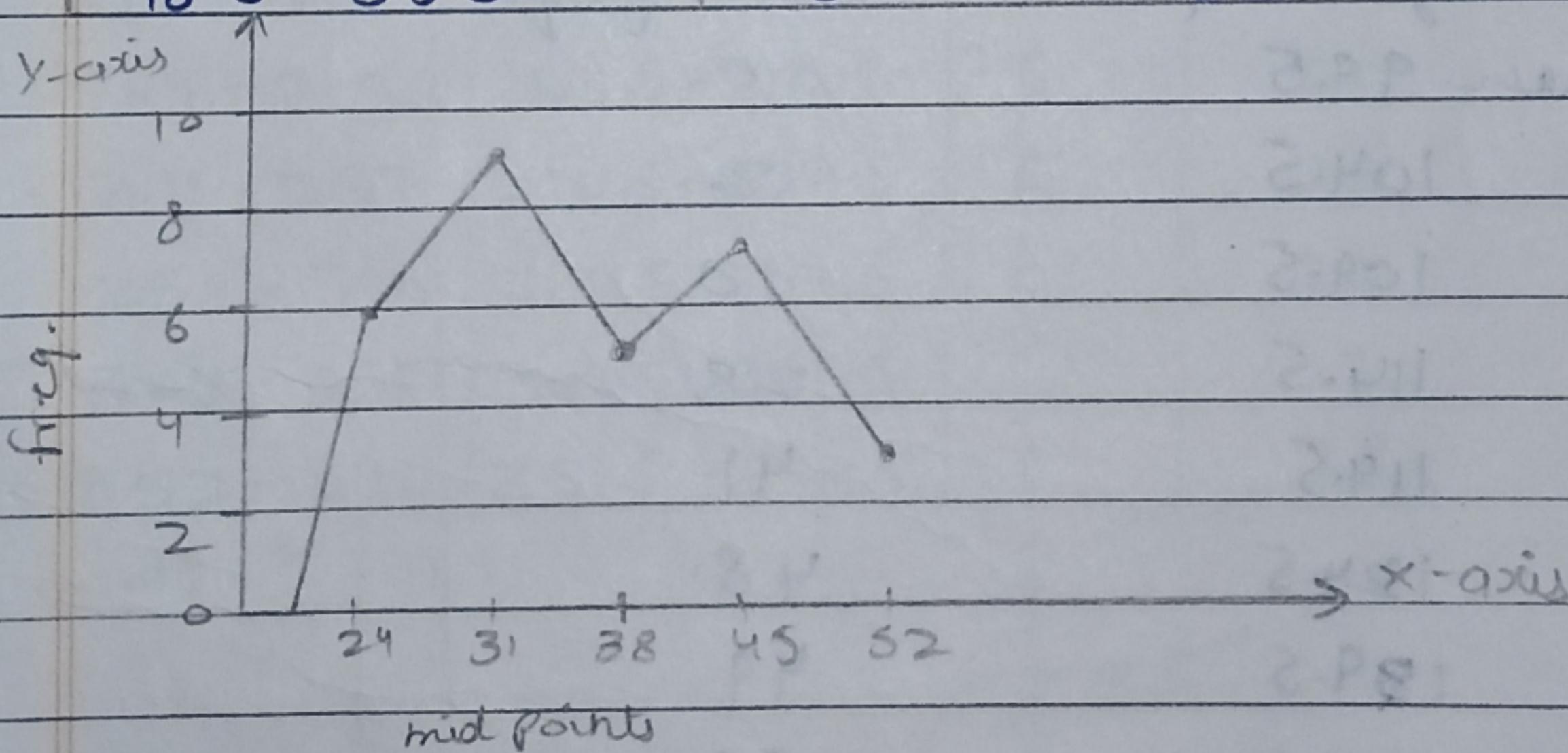
Step II : Draw x-axis and y-axis . x - axis represent mid-points & y - axis represents frequencies.

Step III : Connect adjacent points with the line segment.

(i)	Class boundary	freq.	Mid points
	99.5 - 104.5	2	102
	104.5 - 109.5	8	107
	109.5 - 114.5	18	112
	114.5 - 119.5	13	117
	119.5 - 124.5	7	122
	124.5 - 129.5	1	127
	129.5 - 134.5	1	132

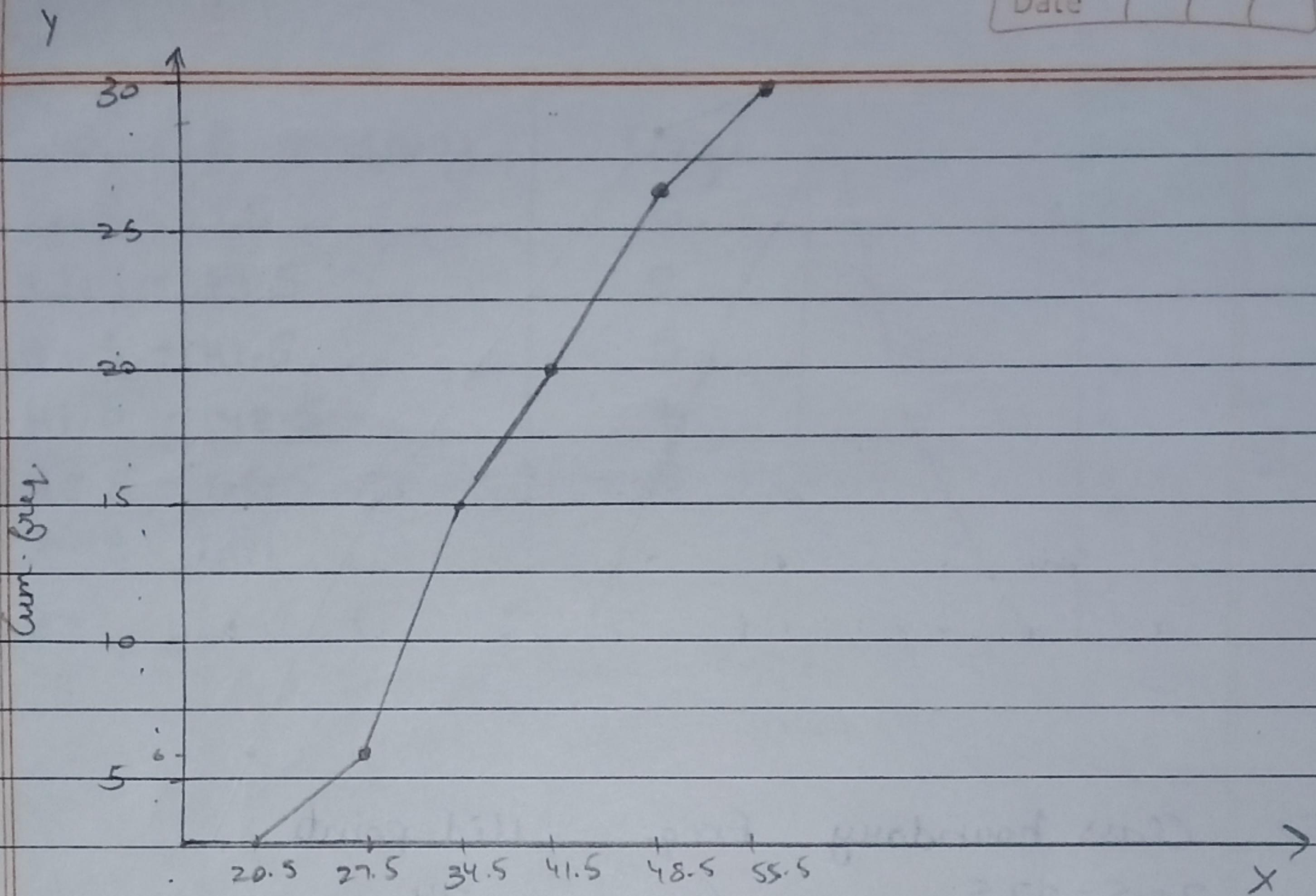


iii	Class boundary	Freq.	Mid-points
	20.5 - 27.5	6	24
	27.5 - 34.5	9	31
	34.5 - 41.5	5	38
	41.5 - 48.5	7	45
	48.5 - 55.5	3	52



- Ogive \rightarrow Cumulative frequency graph

Class boundary	Freq.	Cum. freq.	Less than 20.5 \rightarrow 0
20.5 - 27.5	6	6	Less than 27.5 - 6
27.5 - 34.5	9	15	34.5 - 15
34.5 - 41.5	5	20	41.5 - 20
41.5 - 48.5	7	27	48.5 - 27
48.5 - 55.5	3	30	55.5 - 30



Ogive is a graph that represents cumulative freq. for the classes in a freq. distribution.

(ii) Records high temp.
Less than 99.5

Cum. freq

0

2

10

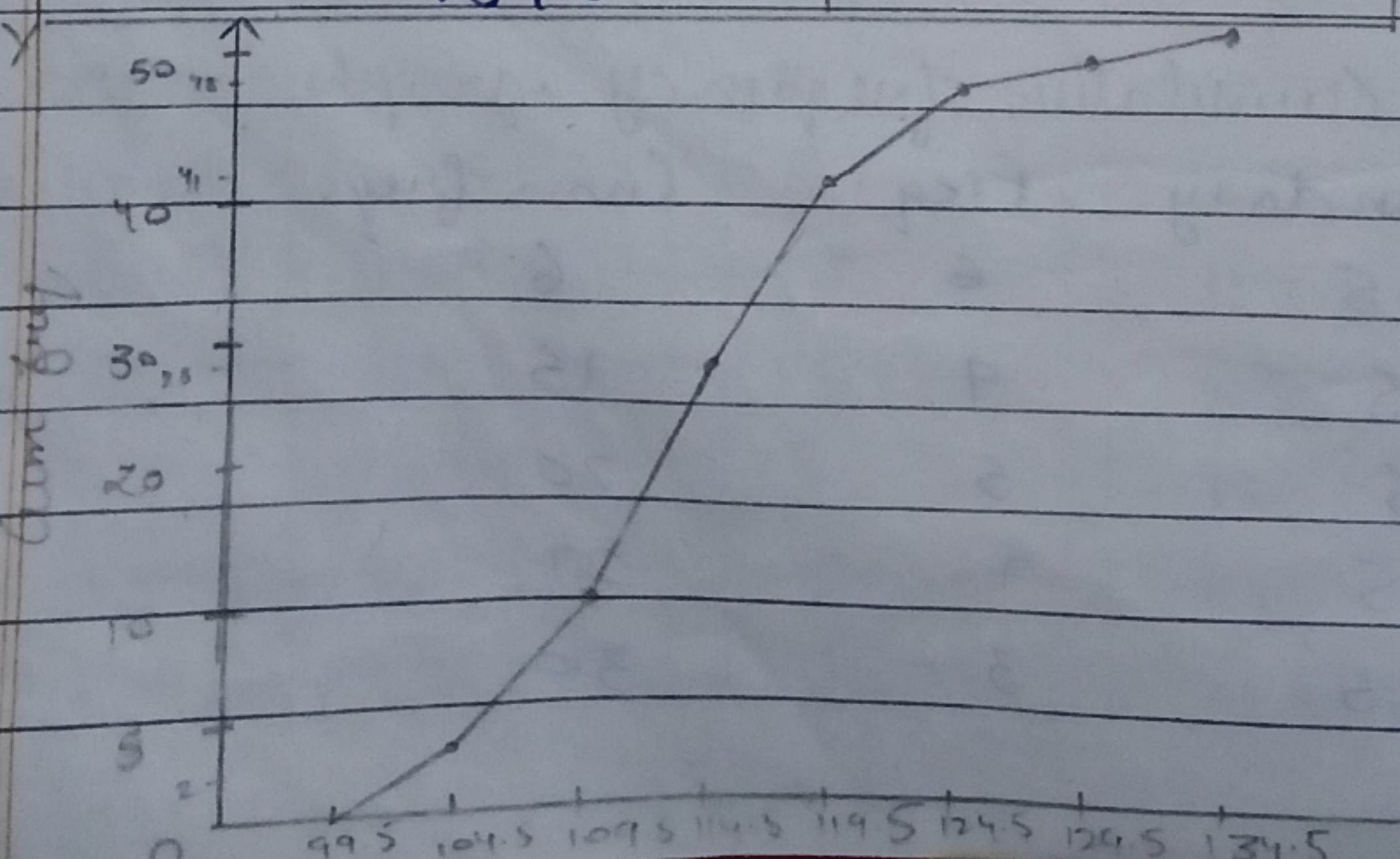
28

41

48

49

50



- The no. of faculty listed for a variety of Private Colleges that offer only Bachelor's degree is listed below. Use these data to construct a freq. distribution with 7 classes, a histogram, a freq. polygon, and an ogive.

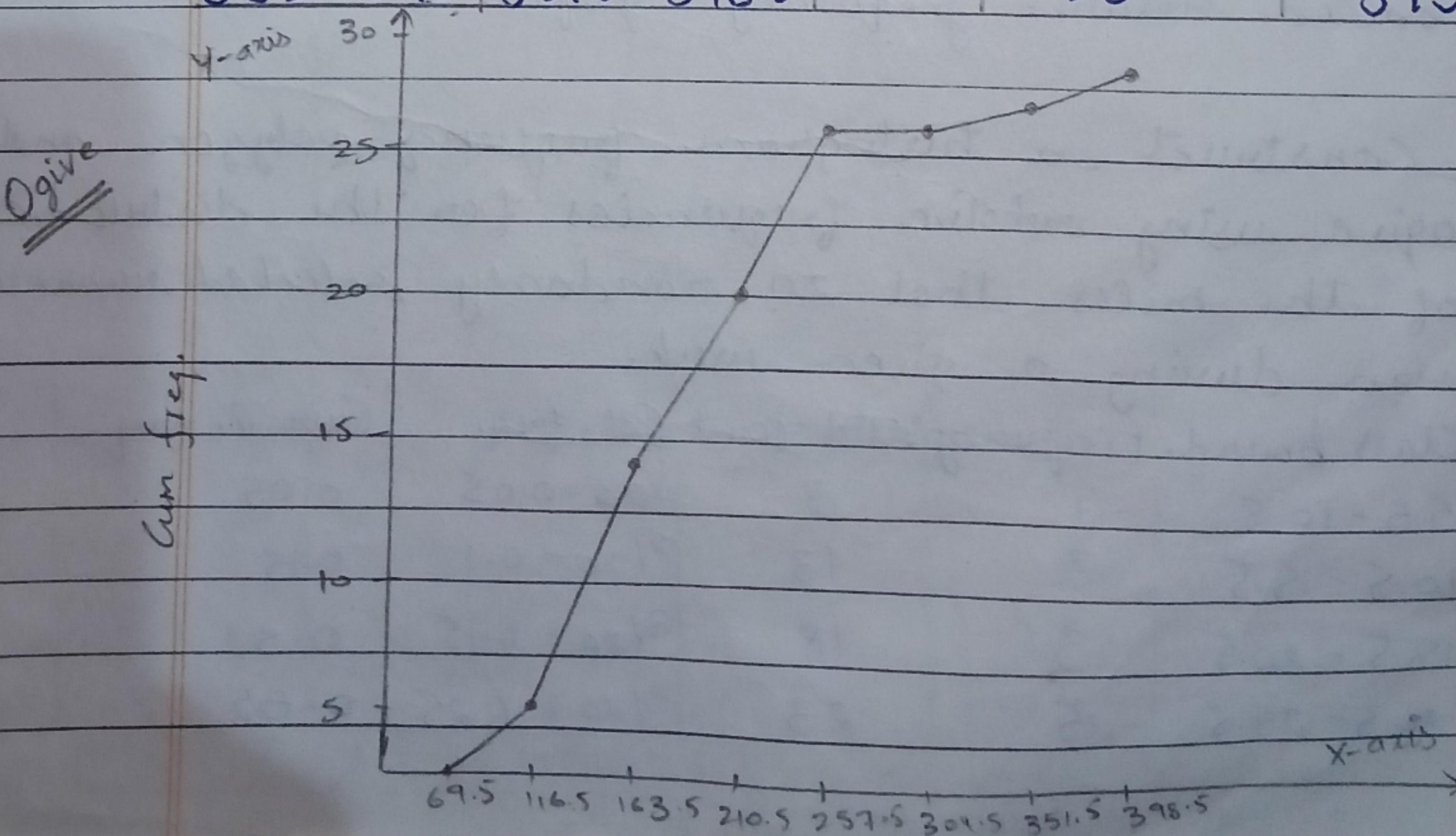
165 221 218 206 138 135 224 209
 70 210 207 154 155 82 120 135
 176 162 225 214 93 389 77 116
 221 161 128 310

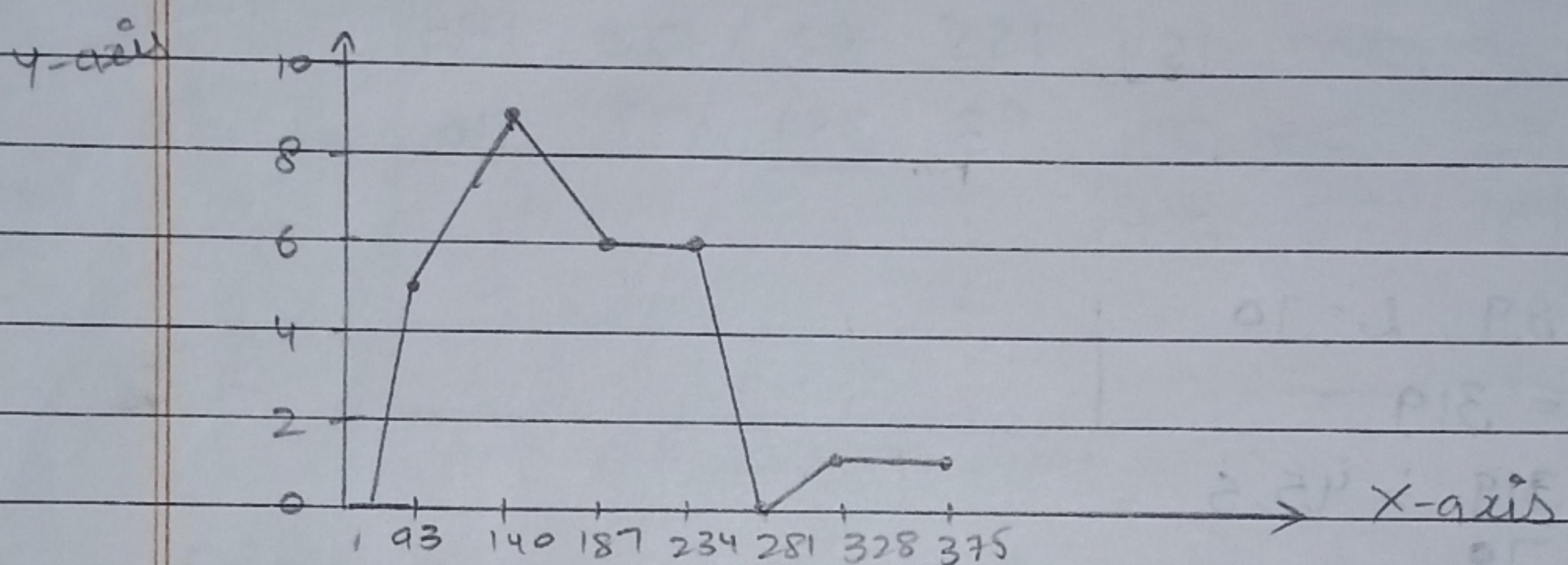
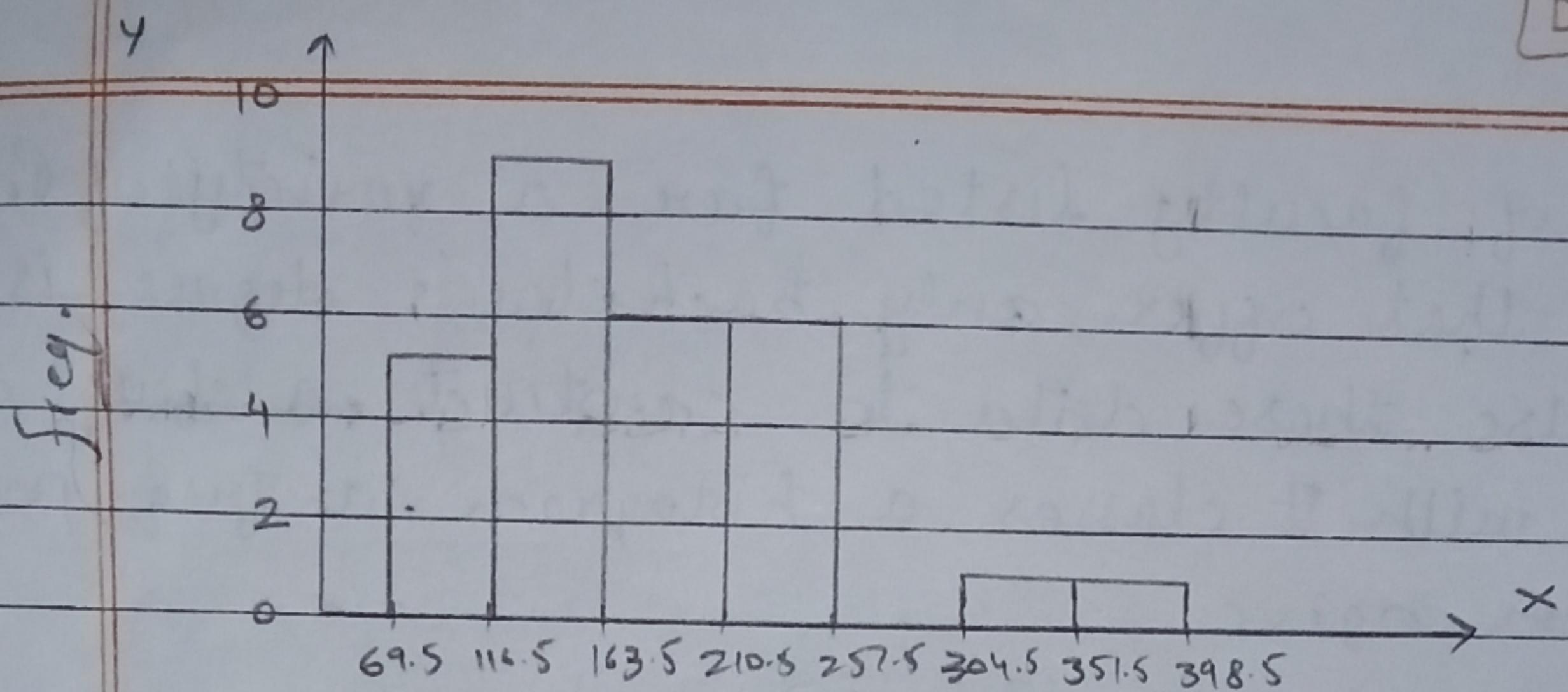
$$\text{I } H = 389, L = 70$$

$$\text{II Range} = 319$$

$$\text{III } W = \frac{319}{7} = 45.5$$

Class int.	Class bound	Freq.	cum.freq.	Mid points
70-116	69.5-116.5	5	5	93
117-163	116.5-163.5	9	14	140
164-210	163.5-210.5	6	20	187
211-257	210.5-257.5	6	26	234
258-304	257.5-304.5	0	26	281
305-351	304.5-351.5	1	27	328
352-398	351.5-398.5	1	28	375





Relative frequency graphs:-

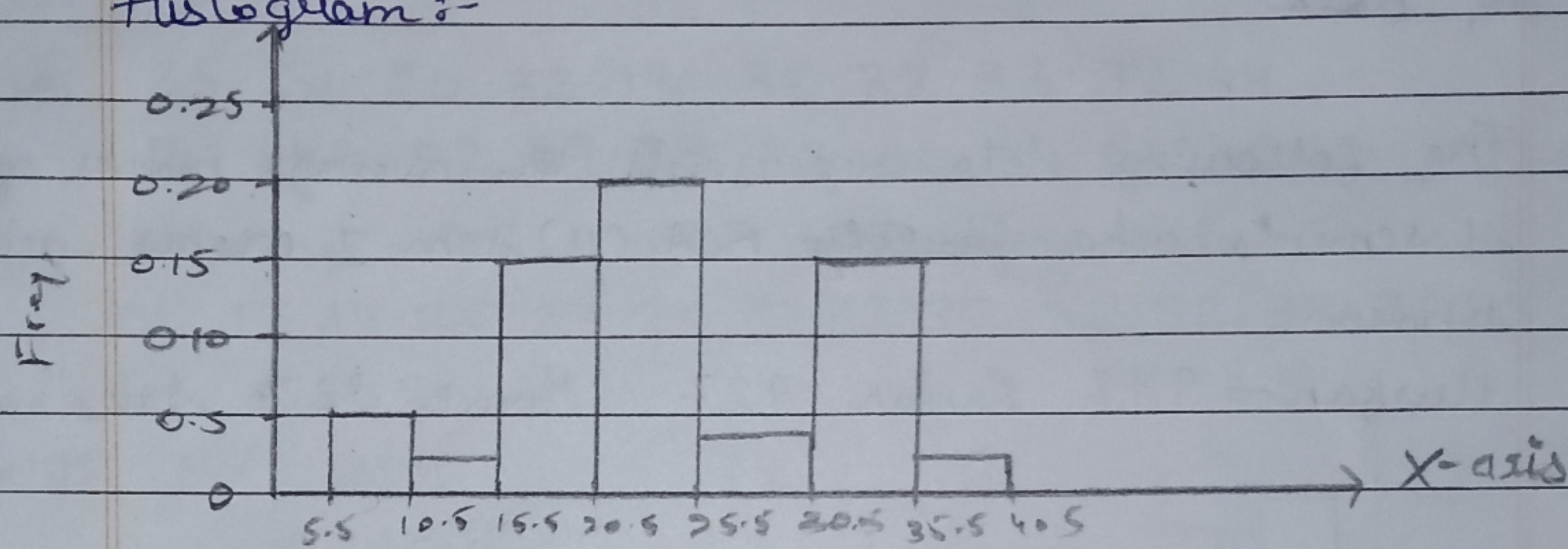
The histogram, the freq polygon, the ogive were constructed by using frequencies in terms of the raw data. The distribution can be converted to distribution using properties instead of raw data as frequencies and these type of graphs are called relative frequency graphs.

Q.1 Construct a histogram, frequency polygon and ogive using relative frequencies for the distribution of the miles that 20 randomly selected runners ran during a given week.

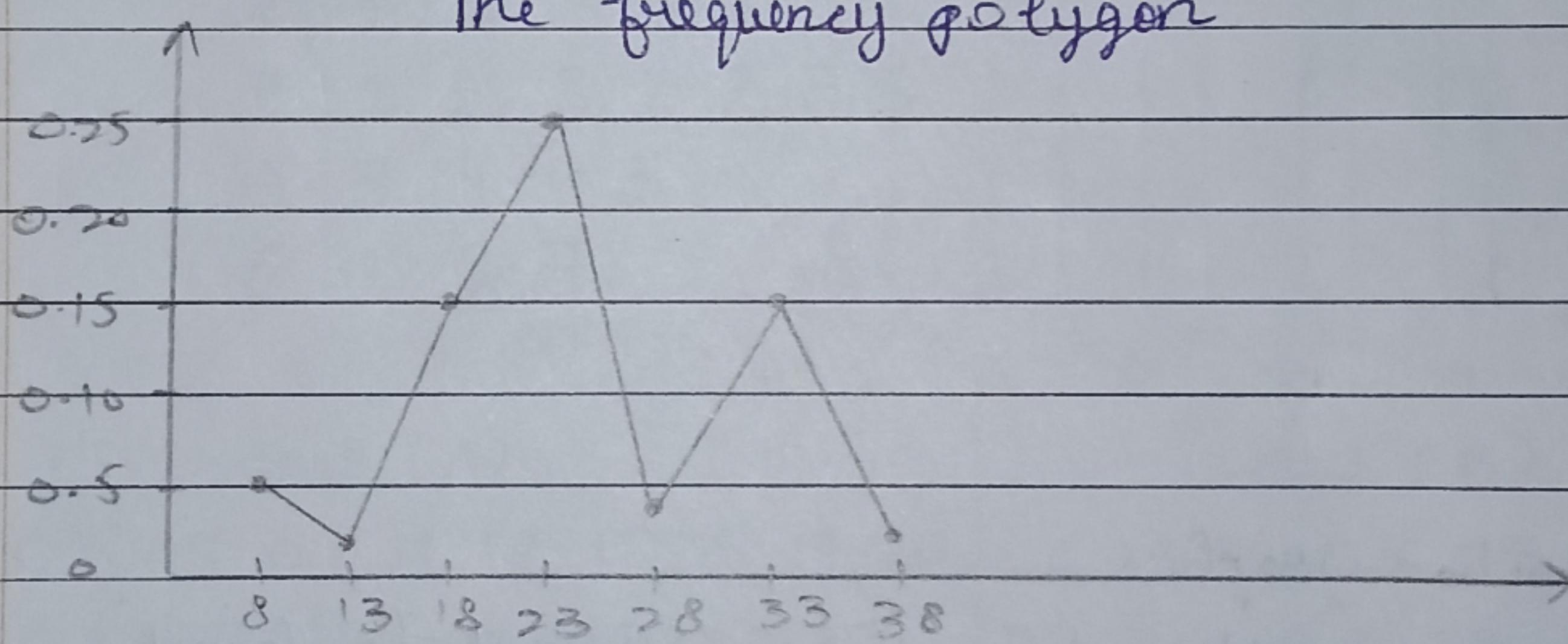
Class bound	Frequency	Mid-point	Rel. freq	Cum R. freq
5.5 - 10.5	1	8	$1/20 = 0.05$	0.05
10.5 - 15.5	2	13	$2/20 = 0.1$	0.15
15.5 - 20.5	3	18	$3/20 = 0.15$	0.30
20.5 - 25.5	5	23	$5/20 = 0.25$	0.55

25.5-30.5	4	28	$4/20 = 0.2$	0.75
30.5-35.5	3	33	$3/20 = 0.15$	0.90
35.5-40.5	2	38	$2/20 = 0.1$	1

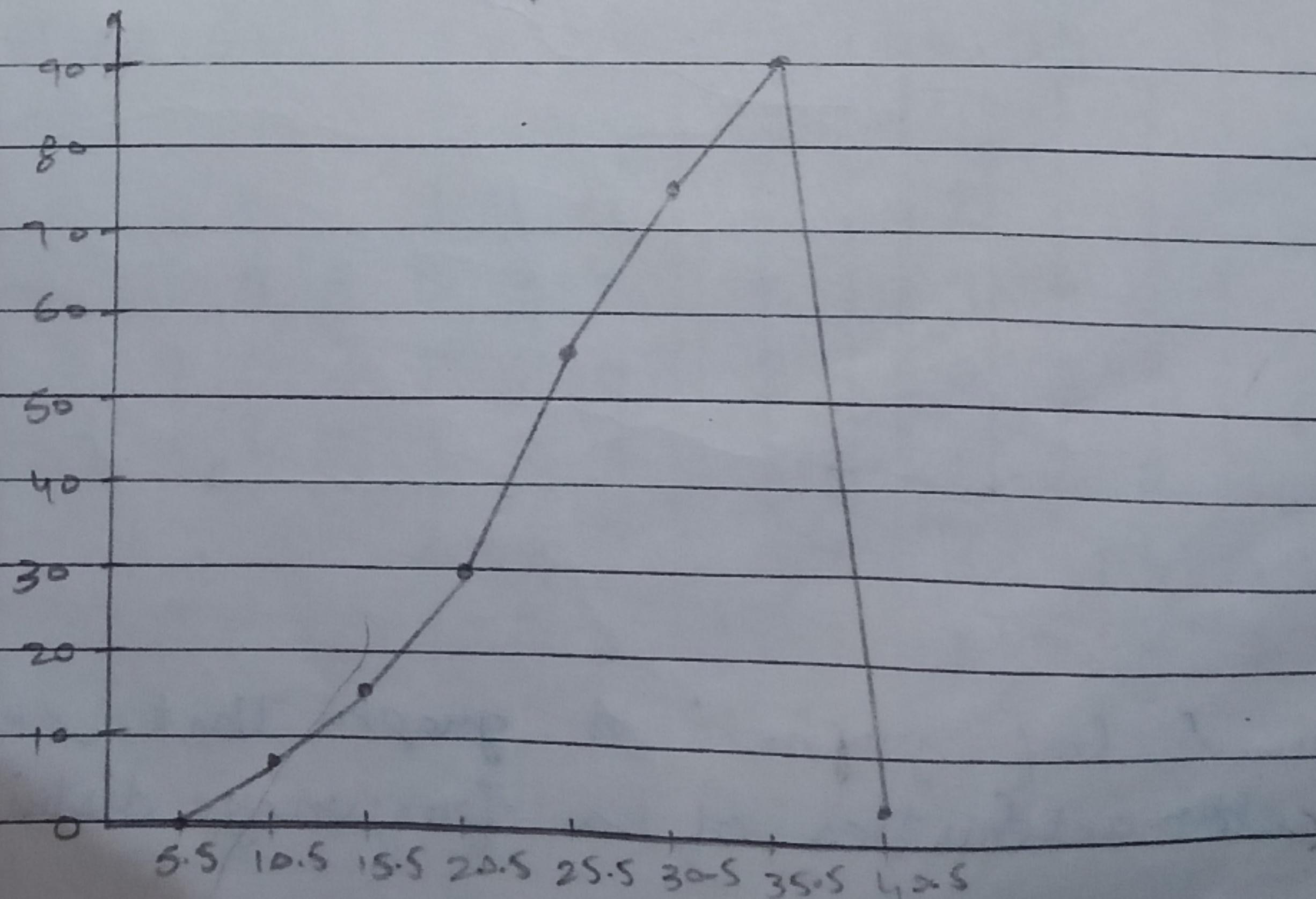
Histogram :-



The frequency polygon



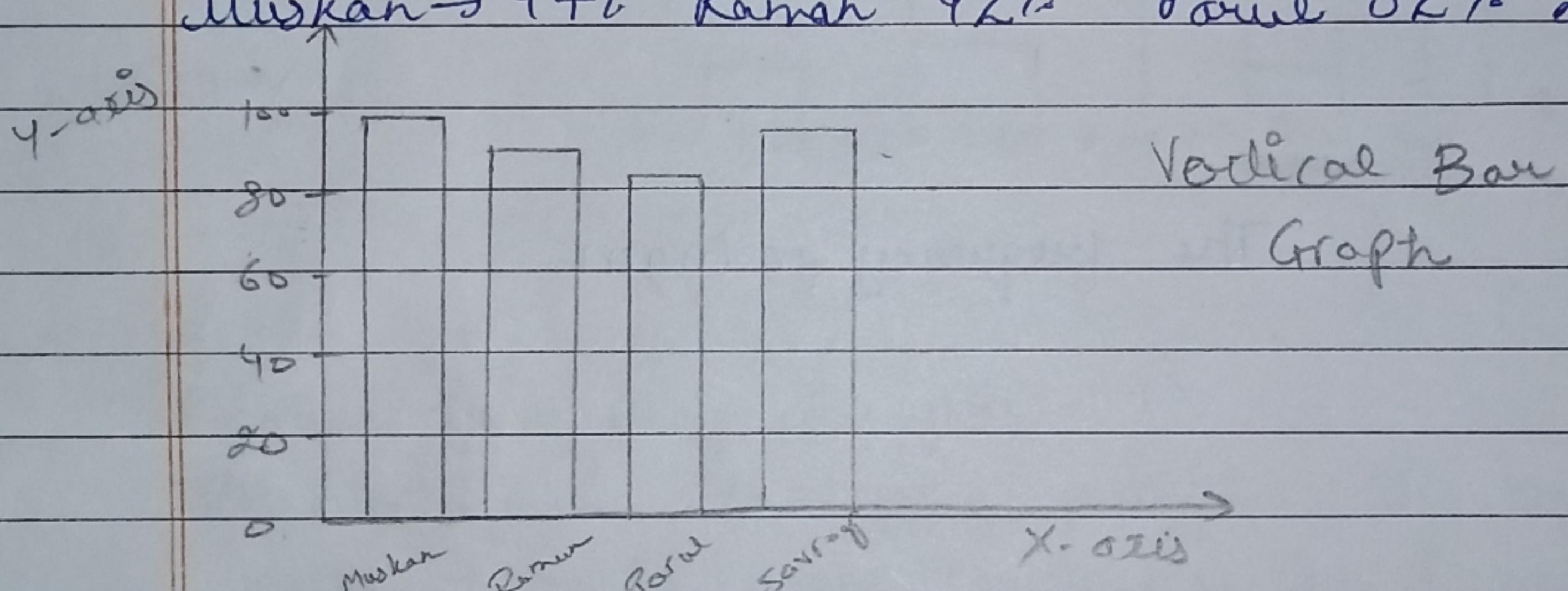
Mid points



- When the data are qualitative or categorical bar graphs can be used to represent the data. A bar graph can be drawn using either vertical or horizontal bars whose heights represents the frequencies of data

The following data represents the average marks of the statements who qualify BCA (H) sem I exams are as under.

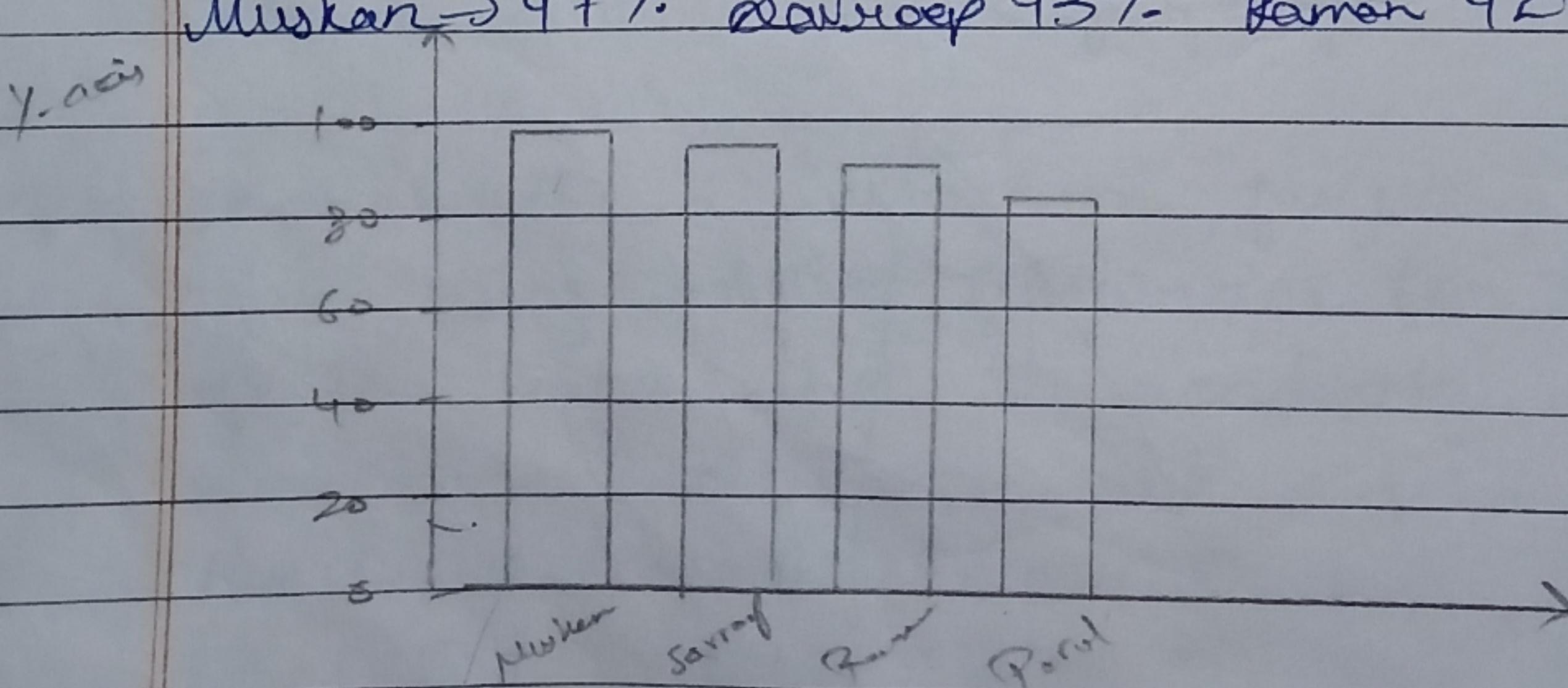
Muskan \rightarrow 97% Raman 92% Powell 82% Savioop 93%



Pareto-graph

Arrange -the data from highest -to lowest.

Muskan \rightarrow 97%. Savioop 93%. Raman 92%. Powell 82%.



Steam & leaf graph :- A graph that combines the characteristics of a frequency distribution

or a histogram is called stem and leaf graph.
 A stem and leaf graph is data plot that uses part of data as a leading digit (stem) & a part of data as a leaf.

Q.no 25 31 20 32 13 36 32 33 32 44
 14 43 02 57 23 32 52 44 51 45.

Ans Arrange data into ascending order

02, 13, 14, 20, 23, 25, 31, 32, 32, 32, 32, 33, 33, 36, 43, 44, 44,
 45, 51, 52, 57

Step II	stem	leaf
	0	2
	1	3 4
	2	0 3 5
	3	1 2 2 2 2 3 6
	4	3 4 4 5
	5	1 2 7

2.

stats (A)

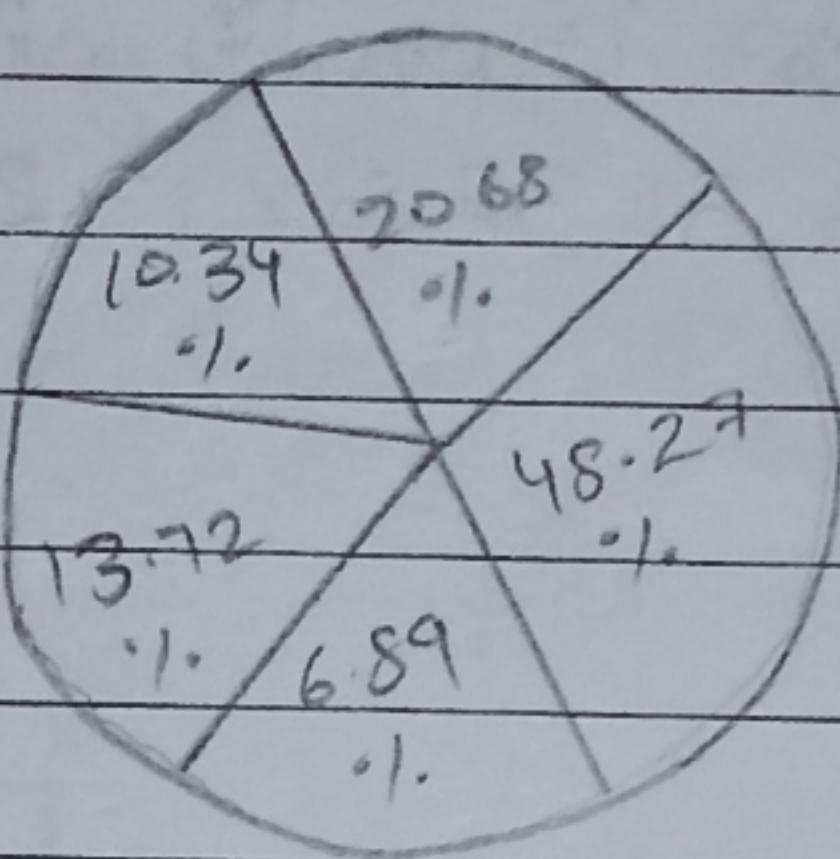
05, 09, 09, 11, 12, 15, 15, 19, 20, 02, 13, 14, 20, 23, 25, 31, 32,
 22, 22, 22, 25, 29, 33, 33, 32, 32, 32, 33, 36, 43, 44, 44,
 33, 33, 34, 35, 37, 40, 41, 42, 45, 51, 52, 57, 57, 57, 58, 58,
 51, 52, 59 58, 59.

stats (B)

stem
9 9 5
9 5 5 2 1
9 5 2 2 2 0
7 5 4 3 3 3 3
2 1 0
9 2 1
5
2
1 3 4
2 0 3 5
1 2 2 2 2 3 6 6
3 4 4 5
1 2 7 7 7 8 8 8 9

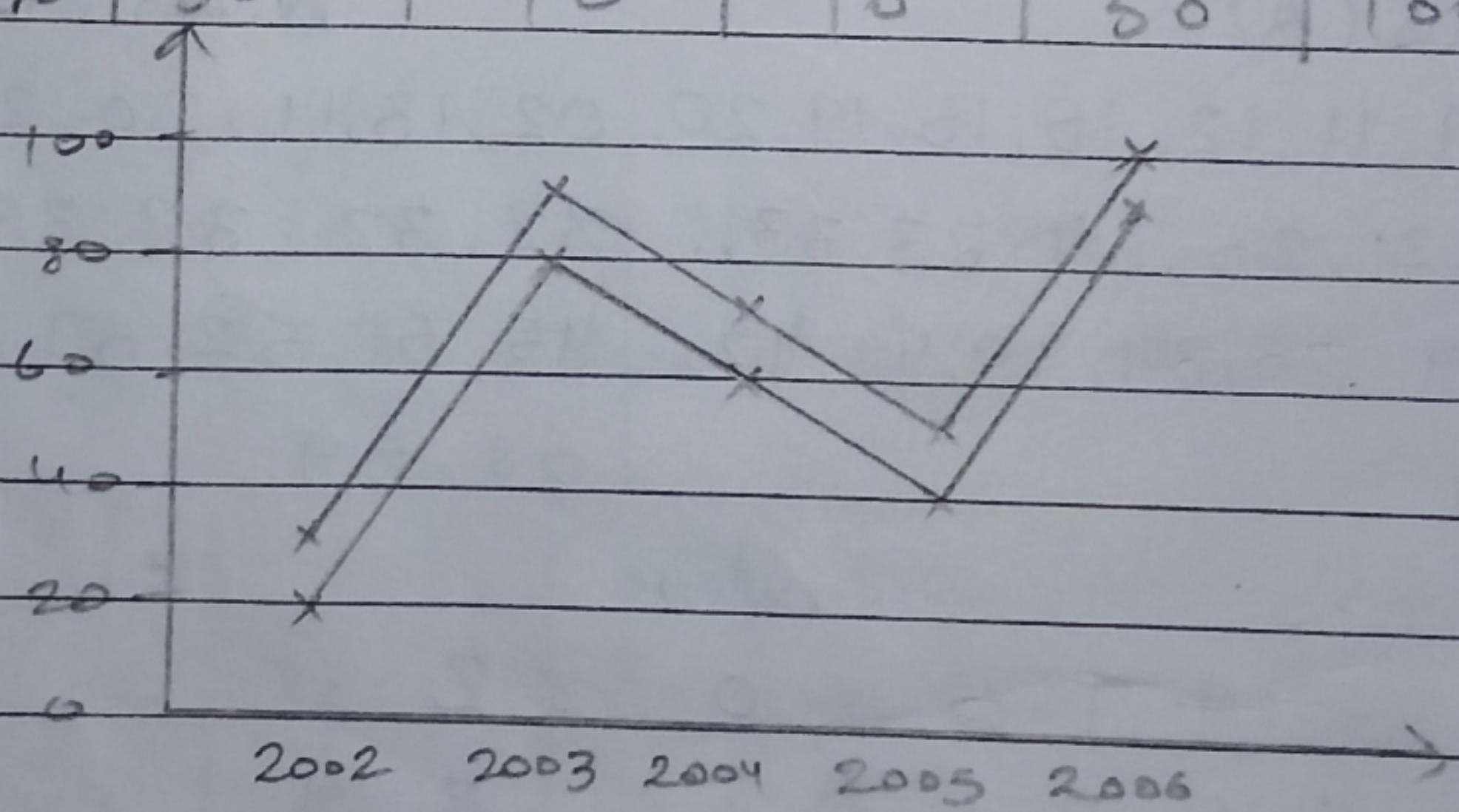
Pie chart :-

Fruits	Frequency	Percentage
Apple	20	$\frac{20}{145} \times 100 = 13.72\%$
Orange	15	10.34%
Banana	70	48.27%
Papaya	30	20.68%
Pineapple	10	6.89%



Time series graph:-

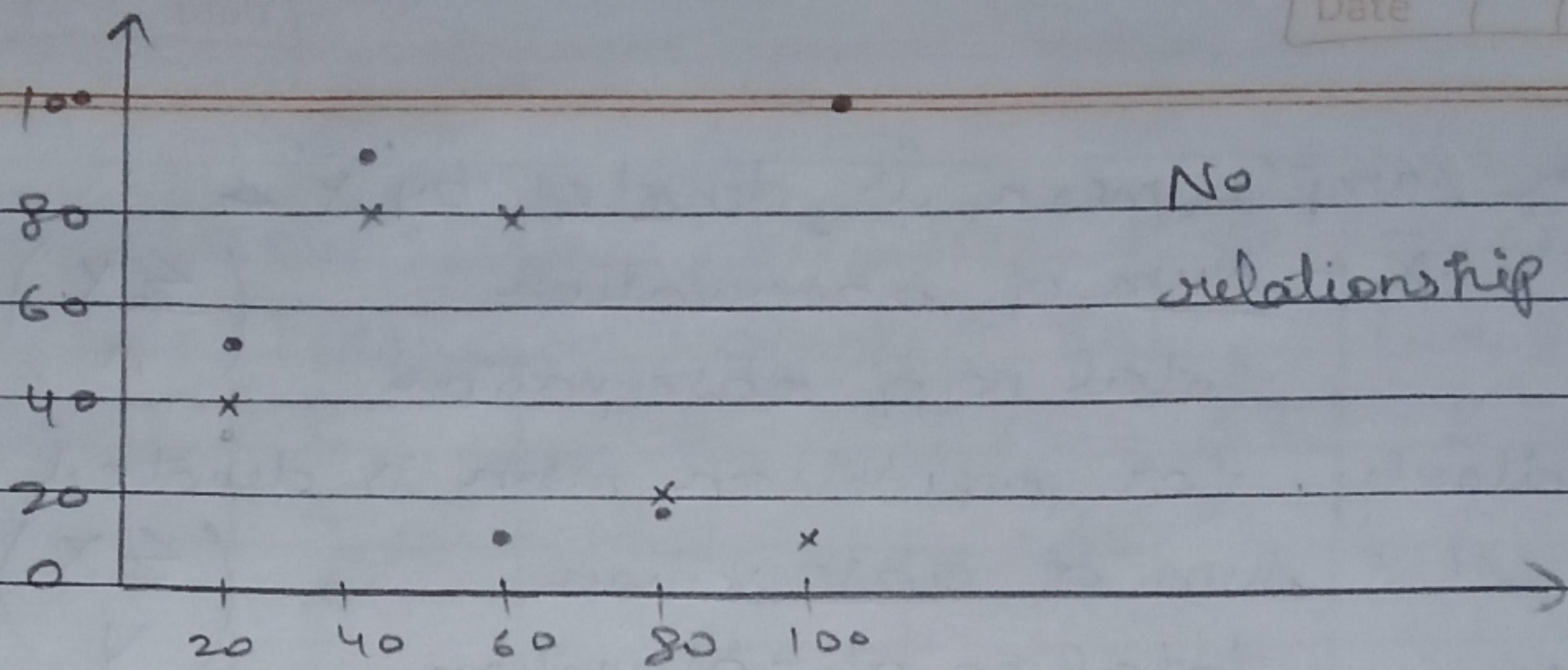
Year	2002	2003	2004	2005	2006
Men	20	80	60	40	90
Women	30	90	70	50	100



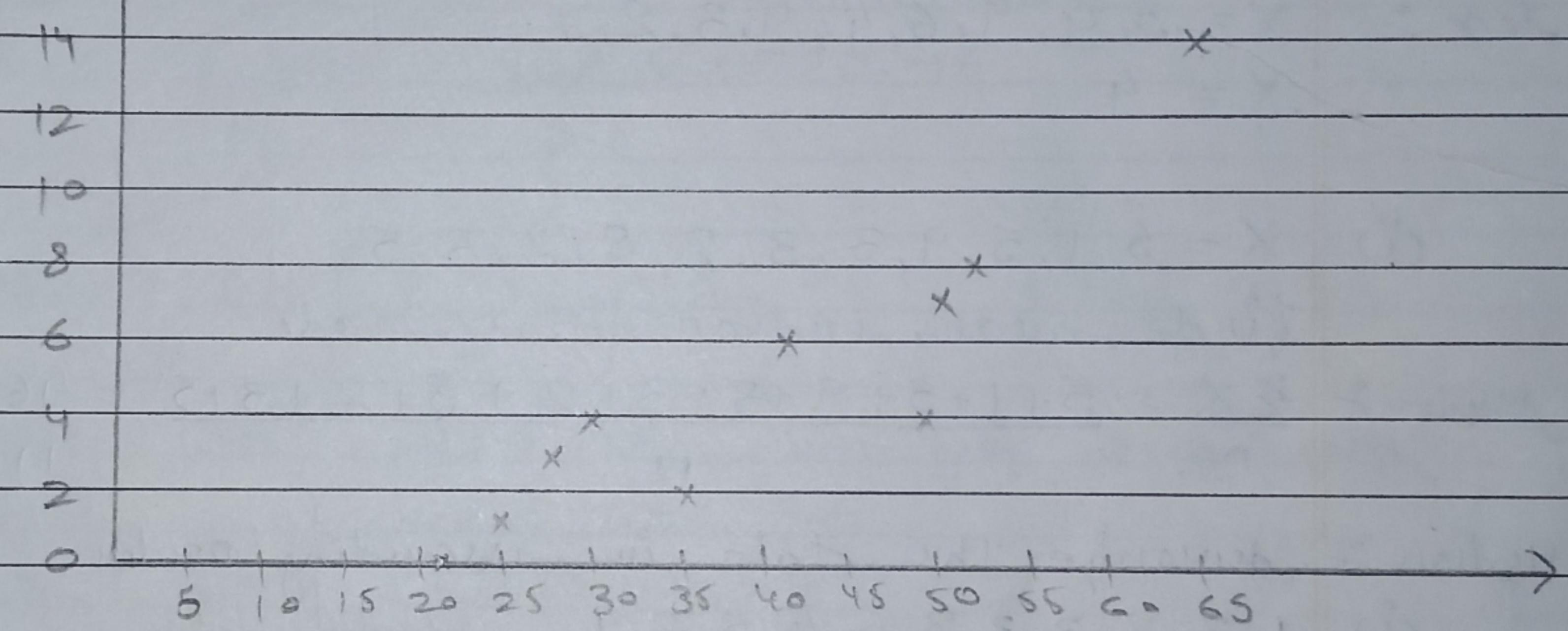
Scatter plot graph

Mens	50	90	10	15	100
Women	40	80	80	20	10

Types :- no relationship, +ve linear relationship, -ve linear relationship, non linear relationship



o	Age (X)	22	30	25	35	65	50	27	53	42	48
	Days (Y)	0	4	1	2	14	7	3	8	6	4



Unit = 3

Data description :-

- * Measure of central tendency
- Mean, Mode, Median for simple data & group data.
- * Measure of dispersion.
- Range, S.D., Variance for population & sample
- * Measure of position.
- Percentiles, Quartiles, Deciles, IQR, Kurtosis.

(i) Mean / Averages :- It can be defined as sum of the observation to the total no. of observations.
 Mean (μ), Sample (\bar{x})