CHAPTER 14
STATISTICS

- **Primary Data:** If an investigator collects data with a specific purpose, then it is called a primary data.
- **Secondary Data:** If a investigator can get a data indirectly from some other source, then the data is called a secondary data.
- **Range** = maximum observation – minimum observation
- **Measures of Central Tendency:**
  - **Mean of ungrouped data:** If $x_1, x_2, x_3, \dots\dots\dots\dots\dots\dots\dots, x_n$ are different observations with respective frequencies $f_1, f_2, f_3, \dots\dots\dots\dots\dots\dots\dots, f_n$, then

    $$\text{Mean} = \frac{1}{N}(x_1f_1 + x_2f_2 + x_3f_3 + \dots\dots\dots\dots\dots\dots\dots + x_nf_n)$$

    $= \frac{1}{N}\sum_{i=1}^{n} x_if_i$, where N= total number of observations

    $= f_1 + f_2 + f_3 + \dots\dots\dots\dots\dots\dots\dots + f_n$
  - **Median of an ungrouped data:**

    The observations are first arranged in ascending or descending order.

    If the number of observations N is odd, then $\left(\frac{N+1}{2}\right)^{th}$ observation is the median.

    If the number of observations N is even, then $\frac{\left(\frac{N}{2}\right)^{th}\text{ term} + \left(\frac{N}{2}+1\right)^{th}\text{ term}}{2}$ is the median.
  - **Mode of an ungrouped data:** The observation with the maximum frequency is the mode.

- **Karl Pearson's Empirical Formula:**

  Mean – Mode = 3(Mean – Median)

Note: In case of a distribution known as the normal distribution, the three measures of central tendency mean, median and mode are coincident.

SOLUTIONS

EXERCISE 14.1

1. State a common experience in which you collect data to get an information.

Ans: Data can be collected personally with a specific purpose or indirectly from other source.

2. Suppose you visited a meteorological centre and collected the record of the maximum temperature on all days of the month July 2007. Is this a primary or a secondary data?

Ans: It is a secondary data.

3. The performance given in grades of 45 students of a class are as follows (A^+, A, B^+, B, C^+, C and D are the 7 grades).

A	**B**	**C^+**	**C**	**C**	**D**	**A^+**	**B**	**B^+**	**B**	**B^+**	**C**	**C^+**	**D**	**A^+**
B^+	**A**	**A^+**	**C**	**C^+**	**A**	**B**	**B^+**	**C**	**D**	**A^+**	**B**	**B^+**	**B^+**	**C**
B	**B^+**	**B**	**B^+**	**B^+**	**A^+**	**C**	**C**	**C^+**	**A**	**A^+**	**B**	**B^+**	**C**	**A**

Construct a frequency distribution to represent the data. Determine the grades which are obtained by the maximum and minimum number of students. Find also the median grade. Which one of the measures of central tendency is not determinable for this non-numerical data?

Solution:

Grades	Tally mark	No. of Students	Cumulative frequency
A^+	𝍸 I	6	6
A	𝍸	5	11
B^+	𝍸 𝍸	10	21
B	𝍸 III	8	29
C^+	IIII	4	33
C	𝍸 IIII	9	42
D	III	3	45
	Total	N = 45	

The grades which are obtained by the maximum and the minimum number of students are B^+ (10 times) and D (3 times) respectively.

Here, N = 45, which is odd.

$$\therefore \text{Median} = \left(\frac{N+1}{2}\right)^{\text{th}} \text{observation} = \left(\frac{45+1}{2}\right)^{\text{th}} \text{observation} = 23^{\text{rd}} \text{ observation} = \text{B}$$

Mean is not determinable for such non-numerical data.

4. What are the mean and median of the first eleven natural numbers? Can you estimate the mode? If so what is it?

Solution:

The first eleven natural numbers are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11.

Mean $= \frac{1+2+3+4+5+6+7+8+9+10+11}{11} = \frac{66}{11} = 6$

Here, N=11, which is odd.

$\therefore$ Median $= \left(\frac{N+1}{2}\right)^{th}$ term $= \left(\frac{11+1}{2}\right)^{th}$ term=6th term =6

Yes, mode can be estimated.

Mean – Mode = 3(Mean – Median) [by Karl Pearson's Empirical formula]

$\Rightarrow$ Mean – 3Mean + 3Median= Mode

$\Rightarrow$Mode = 3Median– 2Mean= $3 \times 6 - 2 \times 6 = 18 - 12 = 6$

5. The following are the weight in kg of 20 pupils in a class

30, 37, 45, 32, 26, 40, 26, 31, 30, 35,
28, 32, 35, 41, 28, 27, 42, 36, 32, 36

Using tally marks form an ungrouped frequency table thereby showing the frequency and cumulative frequency columns. Also find the three measures of central tendency of the data.

Solution:

Weight in Kg (x_i)	Tally Mark	No. of pupils (f_i)	Cumulative Frequency	$x_i . f_i$
26	II	2	2	52
27	I	1	3	27
28	II	2	5	56
30	II	2	7	60
31	I	1	8	31
32	III	3	11	96
35	II	2	13	70
36	II	2	15	72
37	I	1	16	37
40	I	1	17	40
41	I	1	18	41
42	I	1	19	42
45	I	1	20	45
		N=20		$\sum_{i=1}^{13} x_i f_i = 669$

Mean $= \frac{1}{N}\sum_{i=1}^{13} x_i f_i = \frac{1}{20} \times 669 \text{ kg} = 33.45 \text{ kg}$

Here, N =20, which is even.

Median $= \frac{\left(\frac{N}{2}\right)^{th} \text{term} + \left(\frac{N+1}{2}\right)^{th} \text{term}}{2} = \frac{\left(\frac{20}{2}\right)^{th} \text{term} + \left(\frac{20}{2}+1\right)^{th} \text{term}}{2} = \frac{10^{th} \text{term} + 11^{th} \text{term}}{2} = \frac{32+32}{2} \text{kg} = 32 \text{ kg}$

Mode = observation with the maximum frequency= 32 kg

6. Taking 2 as the width of each class construct a continuous grouped frequency table for the sample in article 14.2. Also show the columns of mid-values of the classes and the cumulative frequencies.

Solution:

Class	Mid-value	Tally mark	Frequency	Cumulative frequency
15 – 17	16	I	1	1
17 – 19	18	~~IIII~~ I	6	7
19 – 21	20	~~IIII~~ I	6	13
21 – 23	22	II	2	15
23 – 25	24	II	2	17
25 – 27	26	~~IIII~~ II	7	24
27 – 29	28	~~IIII~~	5	29
29 – 31	30	~~IIII~~	5	34
31 – 33	32	~~IIII~~ I	6	40
33 – 35	34	~~IIII~~	5	45
35 – 37	36		0	45
37 – 39	38		0	45
39 – 41	40	I	1	46
41 – 43	42	I	1	47
43 – 45	44		0	47
45 – 47	46	II	2	49
47 – 49	48		0	49
49 – 51	50		0	49
51 – 53	52	I	1	50

7. For a certain locality, the data of monthly household consumption of electricity measured in units of energy is given below:

Units of Energy Consumed	**1 – 10**	**11 – 20**	**21 – 30**	**31 – 40**	**41 – 50**	**51 – 60**	**61 – 70**	**71 – 80**	**81 – 90**	**91 – 100**	**101 & above**
No. of Consumers	**20**	**35**	**42**	**70**	**100**	**120**	**110**	**65**	**36**	**20**	**10**

The monthly rate of tax is divided into two categories:

Category I: **Rs 3 per unit for households consuming 50 units or less.**

Category II: **Re 1 for every additional unit of energy above 50 units.**

Find the number of consumers paying the first and the second categories of taxes.

Solution:

Units of Energy	Number of Consumers (Frequency)	Cumulative Frequency
1 – 10	20	20
11 – 20	35	55
21 – 30	42	97
31 – 40	70	167
41 – 50	100	267
51 – 60	120	387
61 – 70	110	497
71 – 80	65	562
81 – 90	36	598
91 – 100	20	618
101 & above	10	628

Number of consumers paying the first category of tax = 267

Number of consumers paying the second category of tax = 628 – 267 = 361

8. The following is the survey report of the weekly income of the families of a ward of a municipality.

Income in Rupees	Number of families
80 and above	12050
150 and above	10475
180 and above	8290
250 and above	6470
370 and above	4260
500 and above	2270
750 and above	1180
1000 and above	560
1500 and above	320
2000 and above	150
2500 and above	75
3000 and above	50
3500 and above	24
4000 and above	12
4500 and above	6

From the above data construct a continuous frequency table and also find the cumulative frequency of each class. Further, find the number of families whose weekly income is at the most Rs. 1000.

Solution:

Income in Rupees	Number of families (Frequency)	Cumulative frequency
80 – 150	1575	1575
150 – 180	2185	3760
180 – 250	1820	5580
250 – 370	2210	7790
370 – 500	1990	9780
500 – 750	1090	10870
750 – 1000	620	11490
1000 – 1500	240	11730
1500 – 2000	170	11900
2000 – 2500	75	11975
2500 – 3000	25	12000
3000 – 3500	26	12026
3500 – 4000	12	12038
4000 – 4500	6	12044
4500 and above	6	12050
	12050	

The number of families whose weekly income is at the most ₹ 1000 is 11490

****

Histogram

Representation of a classified data by a graph consisting of rectangles is called the histogram of the data.

Steps to draw the histogram of data:

1. The class limits must be represented along the x-axis on a suitable scale.
2. We construct rectangles whose bases are the class limits and whose areas are proportional to the frequency of the class.

Frequency Polygon

The frequency polygon of a classified data is obtained first by plotting the points whose abscissa are the class marks of the various classes and the ordinates are the frequencies of the corresponding classes and then joining them by line segments.

If we are to draw both histogram and the frequency polygon we get the latter by joining the mid-points of the tops of all rectangles by line segments.

Cumulative Frequency Curve or the Ogive

For a classified data the free-hand smooth curve obtained by joining the points which are plotted by taking the upper limits of the various classes as the abscissa and the cumulative frequencies of the corresponding classes as the ordinates is called the cumulative frequency curve or the ogive of the data.

An ogive is a never descending curve.

SOLUTIONS

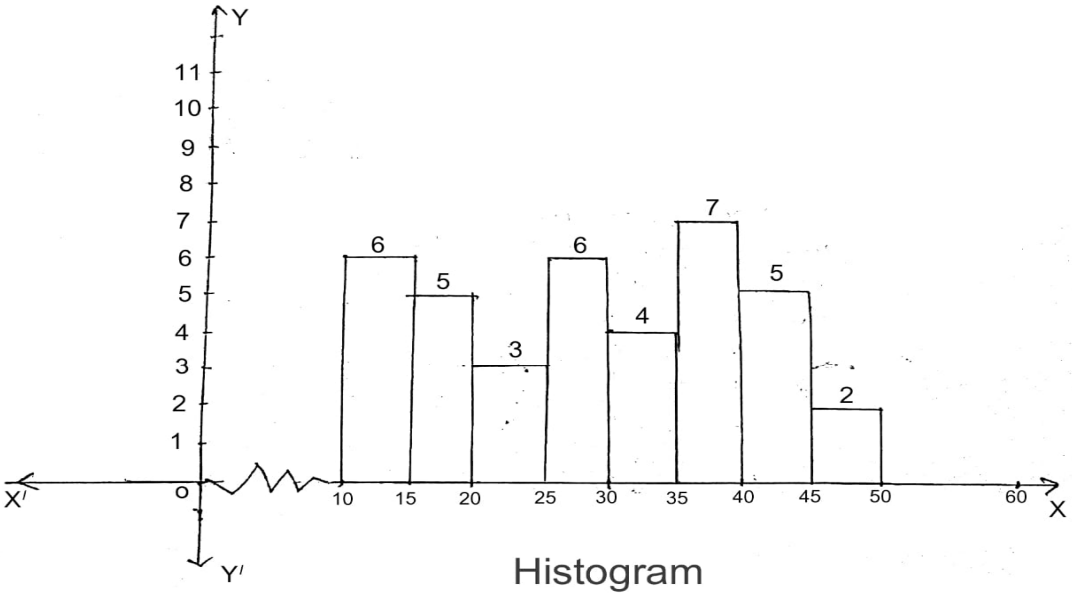
EXERCISE 14.2

1. The following in data of the number of people of a particular village having abnormal blood pressure at different age groups.

Age group	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50
No. of people	6	5	3	6	4	7	5	2

Draw the histogram of this data.

Solution:



Histogram

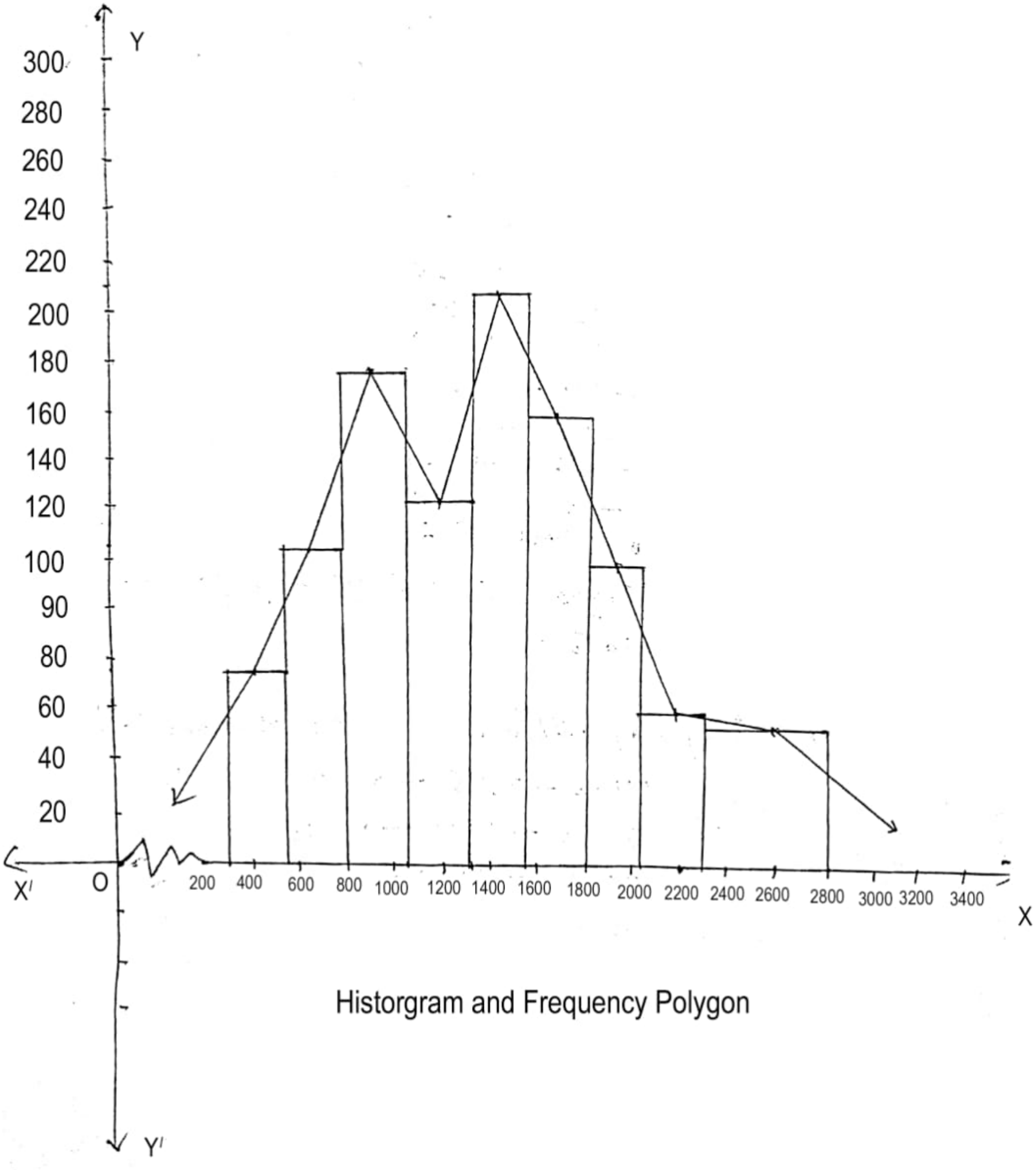
2. The monthly expenditure of the households of a ward of a municipality is given in the following data.

Monthly expenditure	No. of households
501 – 750	75
751 – 1000	125
1001 – 1250	200
1251 – 1500	150
1501 – 1750	230
1751 – 2000	180
2001 – 2250	120
2251 – 2500	60
2501 – 3000	55

Draw the histogram and the frequency polygon of the above data.

Solution:

Monthly expenditure	No. of households
500.5 – 750.5	75
750.5 – 1000.5	125
1000.5 – 1250.5	200
1250.5 – 1500.5	150
1500.5 – 1750.5	230
1750.5 – 2000.5	180
2000.5 – 2250.5	120
2250.5 – 2500.5	60
2500.5 – 3000.5	55



Historgram and Frequency Polygon

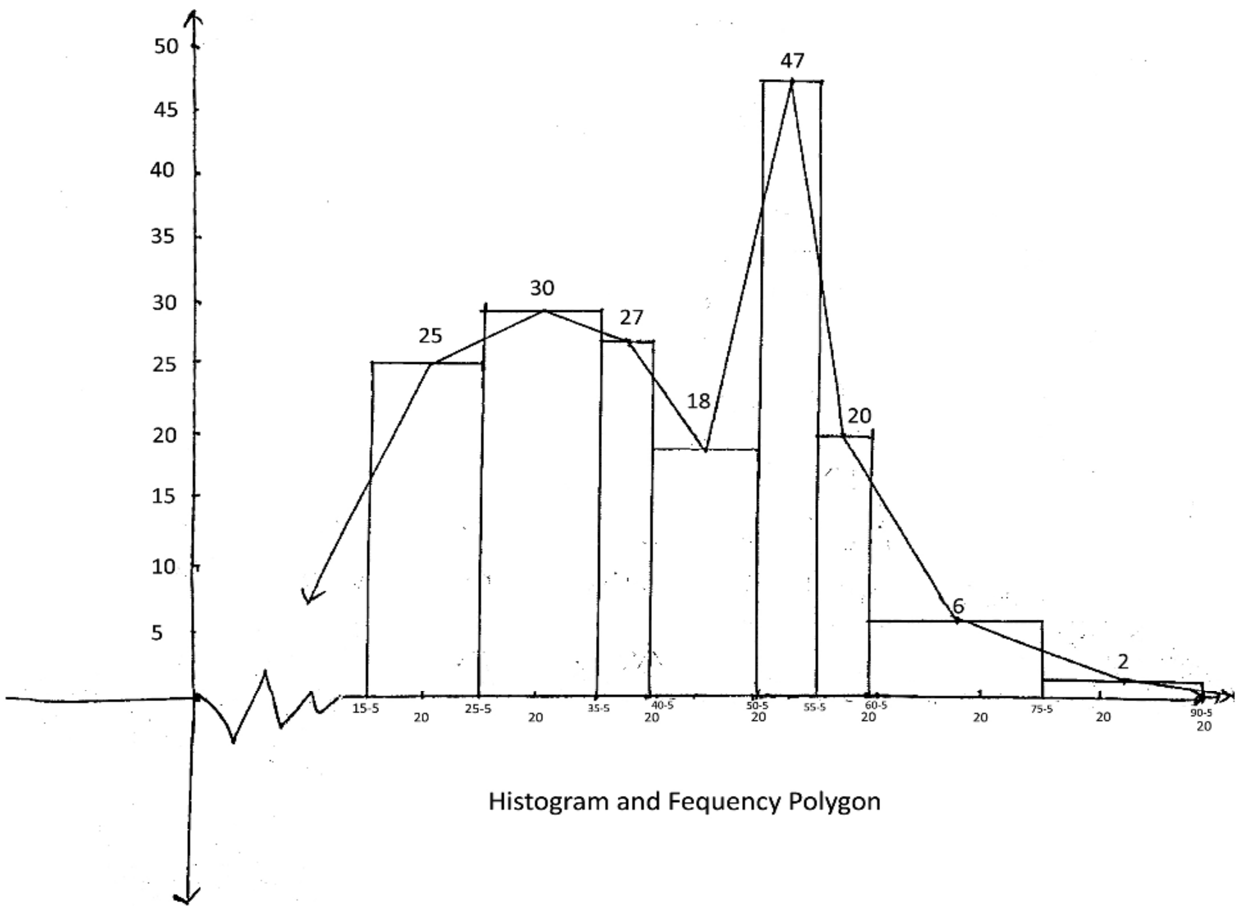
3. The data of occurrence of a particular disease in different age groups of locality is given below:

Age group	15 – 25	26 – 35	36 – 40	41 – 50	51 – 55	56 – 60	61 – 75	76 – 90
No. of people	25	30	27	18	47	20	6	2

Draw the histogram of the above data and also draw the frequency polygon of the same.

Solution:

Age group	No. of people
14.5 – 25.5	25
25.5 – 35.5	30
35.5 – 40.5	27
40.5 – 50.5	18
50.5 – 55.5	47
55.5 – 60.5	20
60.5 – 75.5	6
75.5 – 90.5	2



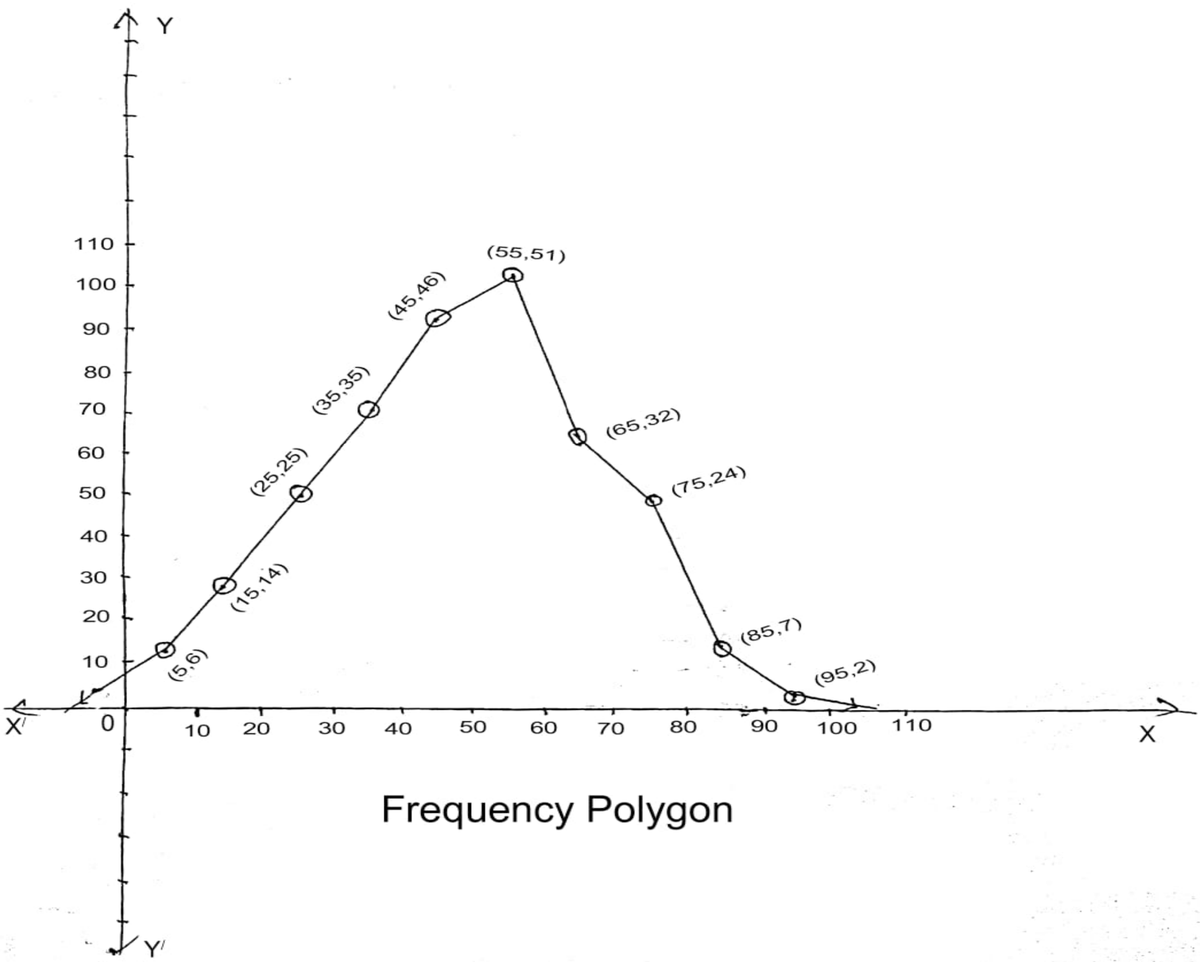
Histogram and Fequency Polygon

4. Draw the frequency polygon of the following data.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of Students	6	14	25	35	46	51	32	24	7	2

Solution:

Marks	Mid-value	No. of Students
0 – 10	5	6
10 – 20	15	14
20 – 30	25	25
30 – 40	35	35
40 – 50	45	46
50 – 60	55	51
60 – 70	65	32
70 – 80	75	24
80 – 90	85	7
90 – 100	95	2



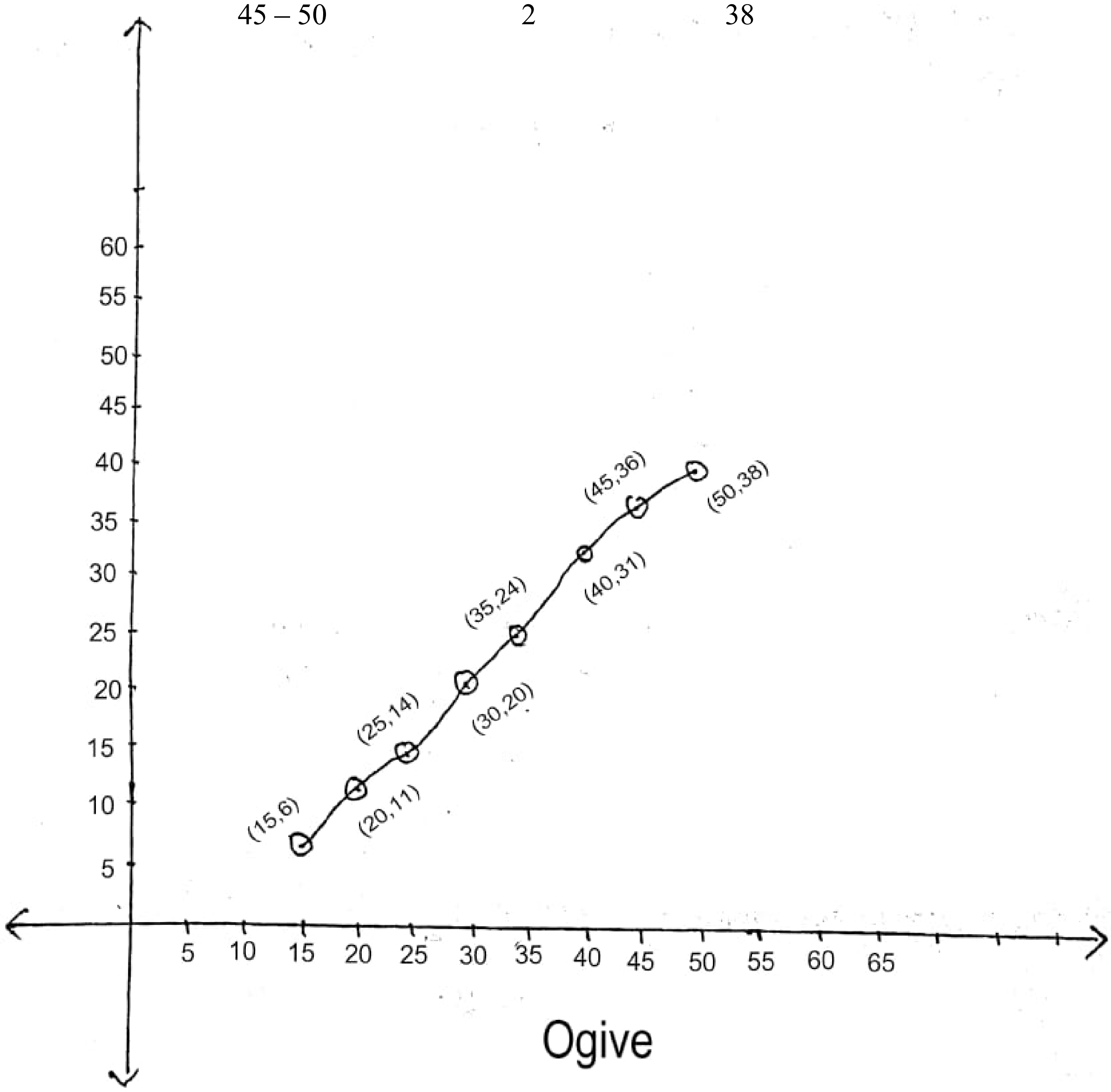
Frequency Polygon

5. Draw the ogive of the data (from Q. no. 1 to 4).

Solution:

(i)

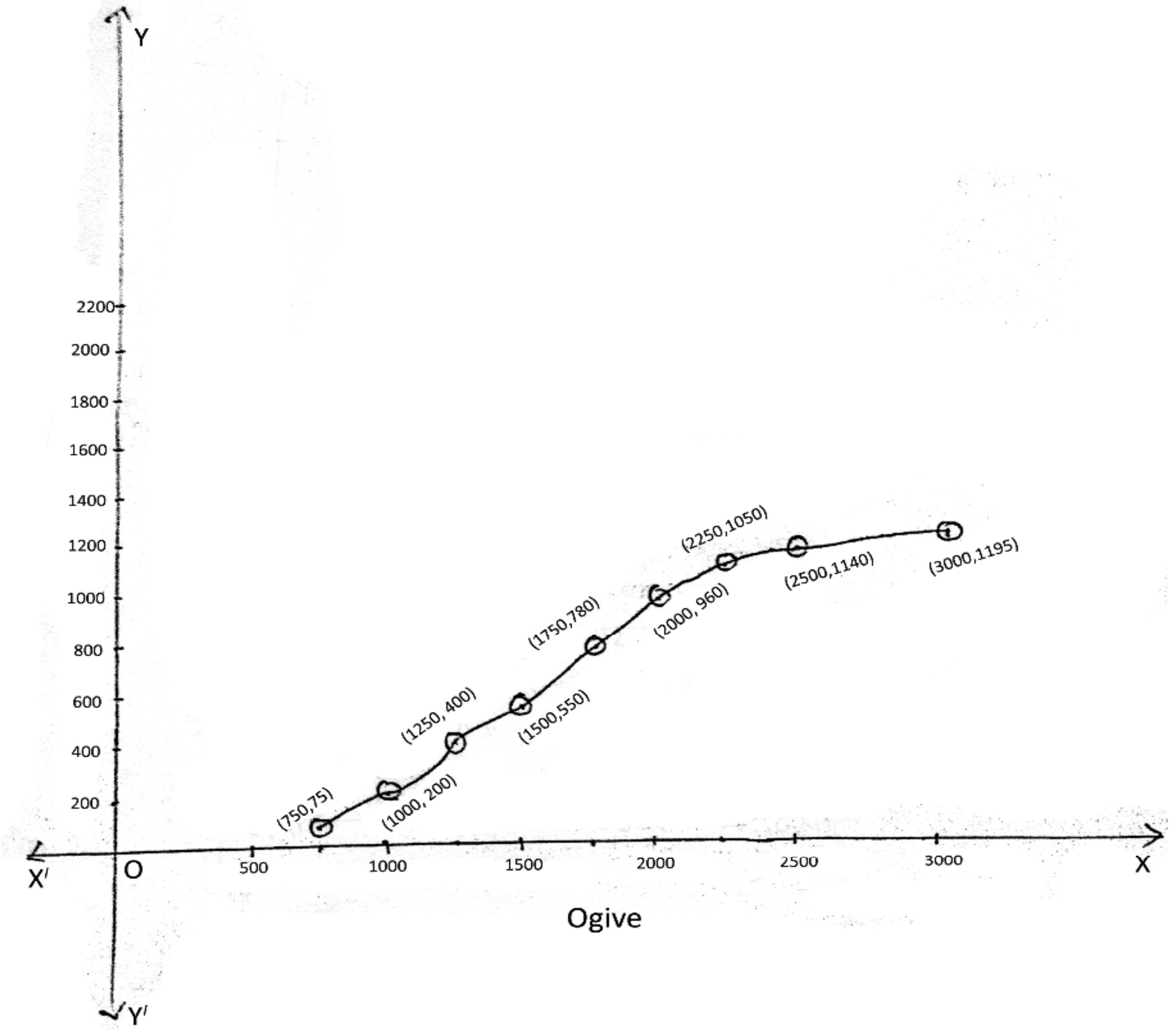
Age group	No. of people (Frequency)	Cumulative Frequency
10 – 15	6	6
15 – 20	5	11
20 – 25	3	14
25 – 30	6	20
30 – 35	4	24
35 – 40	7	31
40 – 45	5	36
45 – 50	2	38



Ogive

(ii)

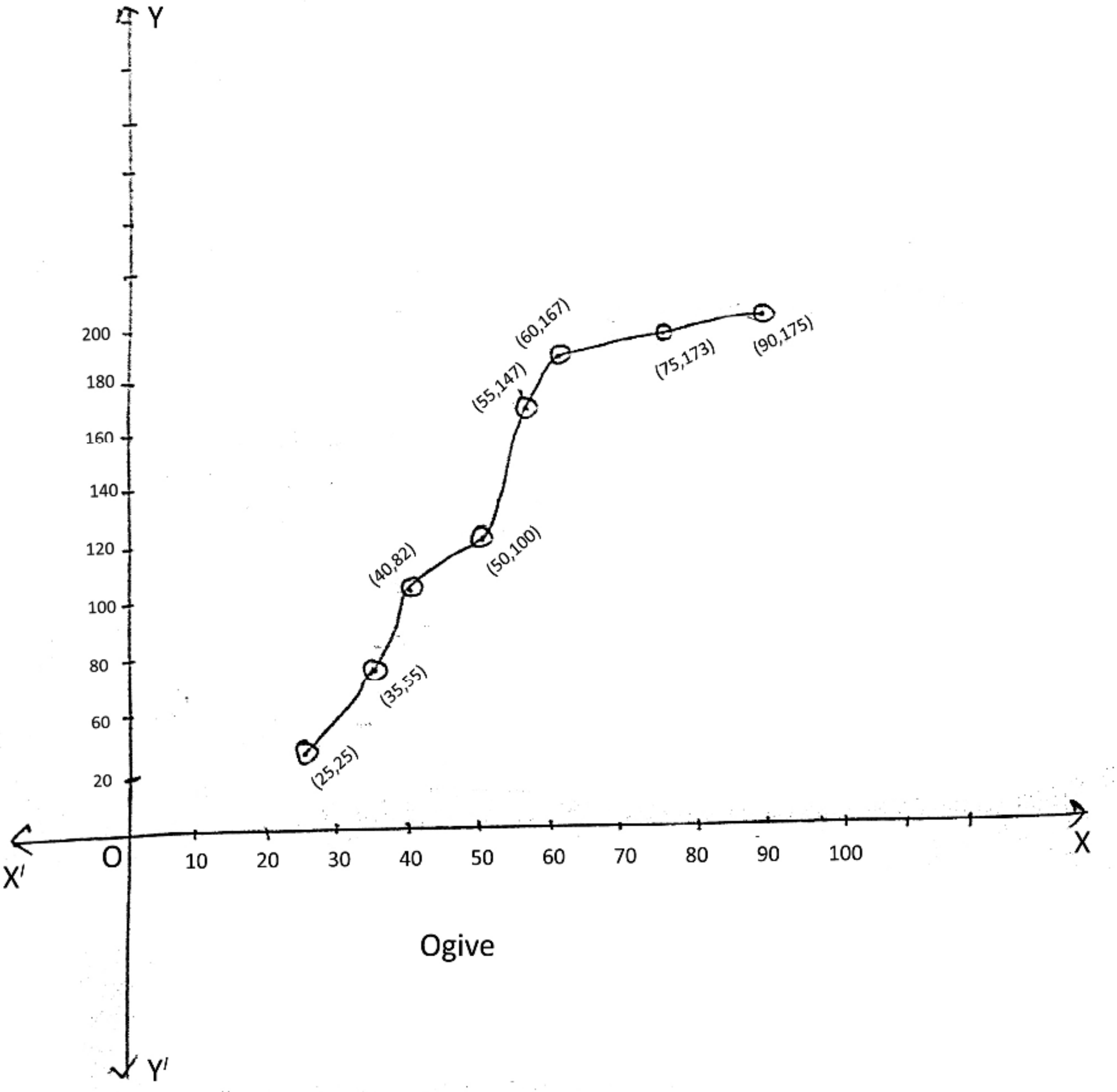
Monthly expenditure	No. of households (Frequency)	Cumulative Frequency
501 – 750	75	75
751 – 1000	125	200
1001 – 1250	200	400
1251 – 1500	150	550
1501 – 1750	230	780
1751 – 2000	180	960
2001 – 2250	120	1080
2251 – 2500	60	1140
2501 – 3000	55	1195



Ogive

(iii)

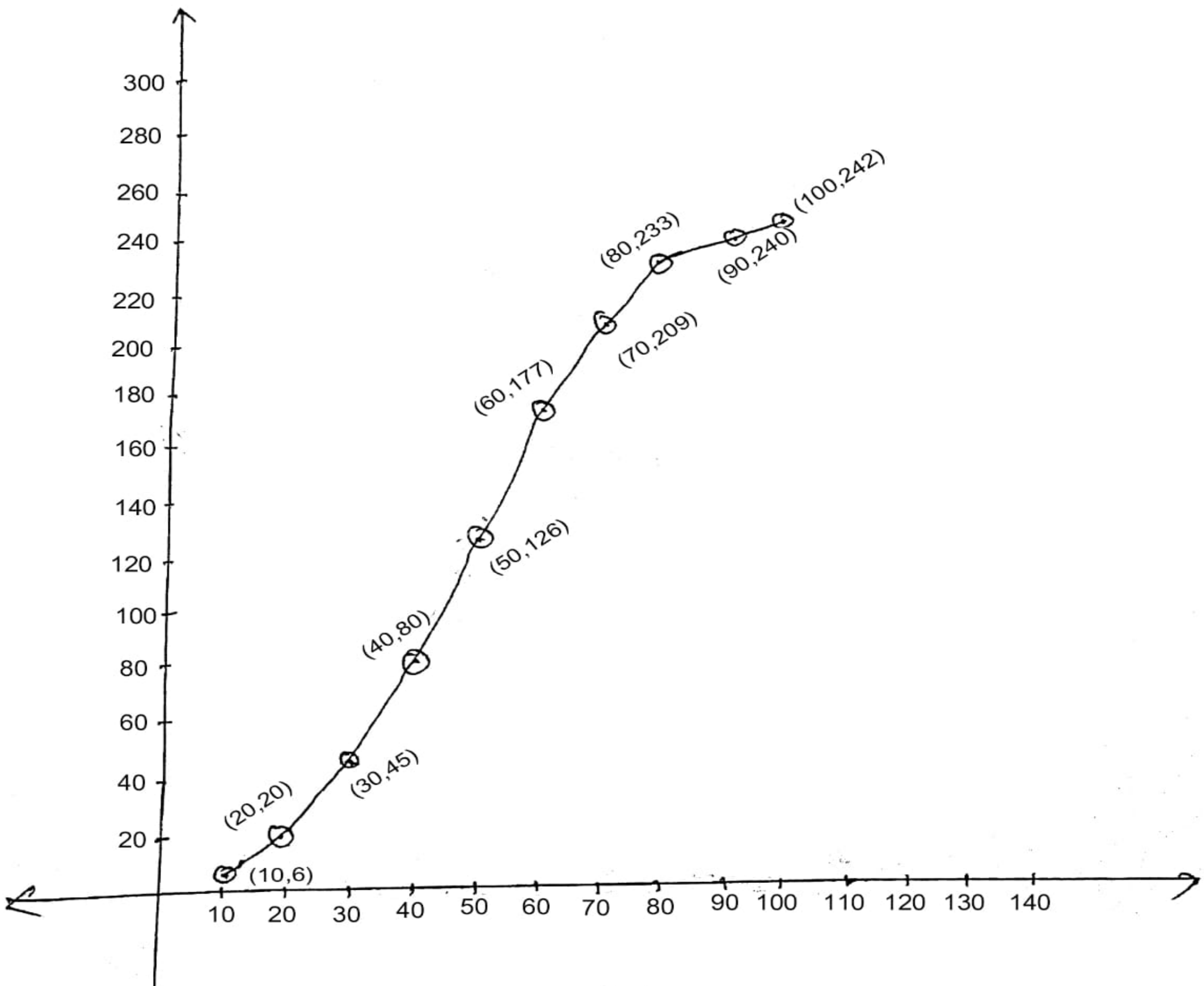
Age group	No. of people (Frequency)	Cumulative Frequency
15 – 25	25	25
26 – 35	30	55
36 – 40	27	82
41 – 50	18	100
51 – 55	47	147
56 – 60	20	167
61 – 75	6	173
76 – 90	2	175



Ogive

(iv)

Marks	No. of Students (Frequency)	Cumulative Frequency
0 – 10	6	6
10 – 20	14	20
20 – 30	25	45
30 – 40	35	80
40 – 50	46	126
50 – 60	51	177
60 – 70	32	209
70 – 80	24	233
80 – 90	7	240
90 – 100	2	242



Ogive