UNIT 5 :
Statistics and Graph Work

18 Statistics

18.1 INTRODUCTION

The word Statistics came into existence towards the middle of the eighteenth century. It seems to have been derived from the Latin word **'Status'** or the Italian word **'Statista'** or the German word **'Statistik'**; each of which means a **'Political State'**.

Formerly, the study of statistics was confined only to the collection of data used for the government purposes; but nowadays, its scope is much wider.

Statistics is used in two different senses — *singular and plural.* When used as *singular,* it refers to the whole subject as a branch of knowledge which deals with statistical principles and methods used in collecting, analysing and interpreting data.

When used as *plural*, it refers to the numerical data, collected in a systematic manner, with some definite object in view, in any field of enquiry.

For example, Statistics of :

(i) marks of students in the ICSE examination.

(ii) number of unemployed persons in different states of India, etc.

In *mathematical statistics,* first of all the numerical facts are collected with some definite object in view and then these facts are organised and analysed to throw light on any sphere of enquiry.

A set of collected numerical facts is called a set of **data**.

For example :

Consider the marks obtained (out of 10) in a class test by 6 students of class X.

Ashok	2	Meeta	7	Beena	8
Hemant	6	Shivani	9		

Here, the set of marks; 2, 8, 9, 7, 6 and 5 is called *a set of data.*

18.2 VARIABLE

A quantity which can vary from one individual to another is called a **variable**.

e.g. height, weight, age, etc.

1. Continuous Variable :

A *variable* which can take any numerical value within a certain range is called a *continuous variable, e.g.*

(i) wages of persons, (ii) heights of children,

(iii) rainfall records of different cities on different days, etc.

2. Discrete (or, discontinuous) **Variable** :

A *variable* which is incapable of taking all possible numerical values is called a *discrete variable.*

e.g. number of children in a family; (the number cannot take any value between 1 and 2, 2 and 3, 3 and 4, etc.).

3. Raw and Arrayed Data

Consider the marks obtained by 20 students in a test of 50 marks as given below :
08, 15, 23, 16, 46, 32, 42, 37, 28, 19, 45, 33, 22, 11, 20, 40, 50, 39, 34 and 49.

The data in the form, as given above, is called **raw** or **ungrouped data**.

If the above data is **arranged in ascending or descending order,** it is called an **arrayed data**.

$\therefore$ Arrayed data in ascending order is :
08, 11, 15, 16, 19, 20, 22, 23, 28, 32, 33, 34, 37, 39, 40, 42, 45, 46, 49 and 50.

And, arrayed data in descending order is :
50, 49, 46, 45, 42, 40, 39, 37, 34, 33, 32, 28, 23, 22, 20, 19, 16, 15, 11 and 08.

18.3 TABULATION OF DATA

Consider *fifteen* children from three families A, B and C. The family A has *five* children out of which *three* are boys and *two* are girls; the family B has *four* children out of which *one* is a boy and *three* are girls; while family C has *six* children out of which *four* are boys and *two* are girls.

These facts can be represented by a table as shown below :

	Boys	Girls	Total
Family A	3	2	5
Family B	1	3	4
Family C	4	2	6
Total	**8**	**7**	**15**

Such a *representation of data* in the form of a table is called **tabulation.**

18.4 FREQUENCY

It is a number, which tells, how many times does a particular data appear in a given set of data.

e.g. consider the following set of data :
1, 2, 0, 3, 2, 1, 5, 4, 3, 2, 1, 2.

In the given set of data, 1 appears 3 times; $\therefore$ frequency of 1 is 3.

Similarly, 2 appears 4 times; therefore, frequency of 2 is 4 and so on.

18.5 FREQUENCY DISTRIBUTION

A tabular arrangement of data showing their corresponding frequencies is called a *frequency distribution.*

The table showing data with their corresponding frequencies is called a *frequency distribution table.*

1. Ungrouped Frequency Distribution :

Consider the following data which gives shoe-sizes of 19 pupils :

4	3	5	6	5	6	5
6	3	6	3	3	6	4
6	6	6	4	5		

In order *to construct a frequency table,* make a table as shown below :

Size	Tally Marks	Frequency
3	IIII	4
4	III	3
5	IIII	4
6	~~IIII~~ III	8

In the first column, write the shoe-sizes in ascending (or, descending) order.

The shoe-sizes of different pupils are recorded in the second column (headed with Tally Marks) by marking a short vertical line called **a stroke.**

The marking of strokes is done as follows :

(i) The first shoe-size in the given data is 4, so make a stroke in the **Tally Marks Column** opposite to size 4.

(ii) The next shoe-size is 3, so make a stroke opposite to size 3 and so on.

(iii) When four strokes are made opposite to any particular shoe-size, don't make the fifth stroke in the same way but make a stroke across the first four. This gives a bundle of five strokes. The next stroke starts a new bundle.

(iv) When marking of the strokes is complete, count the strokes against each size and write in the column headed as : **Frequency.**

The table obtained above shows *ungrouped frequency distribution*.

2. Grouped Frequency Distribution :

Following are the examples of grouped frequency distributions :

1 Given below are the marks obtained by 32 students in an examination :

29	23	30	40	11	01	15	35
40	03	12	18	30	24	25	29
31	32	25	22	27	12	13	02
07	09	19	13	32	39	25	03

Taking class intervals 1 – 10, 11 – 20, ..., 31 – 40, make a frequency table for the above distribution.

Solution :

The frequency table for the given distribution is :

Marks	Tally Marks	Frequency
1 – 10	𝍸 |	6
11 – 20	𝍸 |||	8
21 – 30	𝍸 𝍸 |	11
31 – 40	𝍸 ||	7
		Total 32

Here,
(i) if $x \in 1 - 10$
$\Rightarrow$ $1 \leq x \quad 10$
(ii) if $x \in 11 - 20$
$\Rightarrow$ $11 \leq x \quad 20$
and so on

2 Given below are the marks obtained by 24 students in an examination :

18	17	16	24	25	19	41	22
32	42	44	21	43	26	28	40
29	30	37	27	49	27	34	31

Taking class intervals 10 – 20, 20 – 30, ..., 50 – 60, make a frequency table for the above distribution.

Solution :

The frequency table for the given distribution is :

Marks	Tally Marks	Frequency
10 – 20	||||	4
20 – 30	𝍸 ||||	9
30 – 40	𝍸	5
40 – 50	𝍸 |	6
		Total 24

Here,
(i) if $x \in 10 - 20$
$\Rightarrow$ $10 \leq x < 20$
(ii) if $x \in 20 - 30$
$\Rightarrow$ $20 \leq x < 30$
and so on

In this frequency distribution, the marks 40 are included in the class interval 40 – 50 and not in 30 – 40. Similarly, marks 30 are included in the class interval 30 – 40 and not in 20 – 30.

18.6 TYPES OF FREQUENCY DISTRIBUTIONS

In general, there are two types of frequency distributions :

(i) Inclusive (ii) Exclusive.

In an inclusive distribution, the upper limit of one class does not coincide with the lower limit of the next class. (refer to solution of example 1). In example 1, let a mark 'x' belong to the class interval 1 – 10; then it means that $1 \leq x \leq 10$.

In an exclusive distribution, the upper limit of one class coincides with the lower limit of the next class (refer to solution of example 2). Let a mark 'x' belongs to the class interval 10 – 20; then it means $10 \le x < 20$.

18.7 CLASS INTERVALS AND CLASS LIMITS

In case of an exclusive frequency distribution (refer to frequency table of example 2), 10 – 20 is called *class interval* which is bounded by two numbers 10 and 20. These numbers are called *class-limits;* the *smaller number* 10 is called the *lower class limit* and the *larger number* 20 is called the *upper class limit.*

Similarly, 20 – 30 is a *class interval* whose *lower class limit* is 20 and *upper class limit* is 30.

In case of an inclusive frequency distribution (refer to frequency table of example 1), 1 – 10 is a class-interval whose *lower limit* is 1 and *upper limit* is 10.

To ensure continuity and to get correct class limits, exclusive distribution should be adopted. To convert inclusive class intervals into exclusive, the following adjustment should be made.

1. Adjustment :

Find the difference between the upper limit of one class and the lower limit of the next class. Divide this difference by 2. The value so obtained is called the *adjustment factor*.

Subtract the adjustment factor from all the lower limits and add it to all the upper limits.

In example 1, the adjustment factor is $\frac{11-10}{2} = 0{\cdot}5$.

The adjusted class would then be as follows :

Marks before adjustment	Marks after adjustment	Frequency
1 – 10	0·5 – 10·5	6
11 – 20	10·5 – 20·5	8
21 – 30	20·5 – 30·5	11
31 – 40	30·5 – 40·5	7

The class-limits obtained after adjustment are called the actual (or true) class-limits.

The actual (or true) class limits are also called **class-boundaries.**

2. Class-Size : The difference between the actual lower limit and actual upper limit of a class interval is called its *class-size*.

In example 1, the class-size = 10·5 – 0·5 = 10

In example 2, the class-size = 20 – 10 = 10 and so on.

3. Class-Mark : The class-mark of the class interval is the value midway between its actual lower limit and actual upper limit.

Thus, class-mark of 0·5 – 10·5 $= \frac{0{\cdot}5 + 10{\cdot}5}{2} = 5{\cdot}5$

Class-mark of 10 – 20 $= \frac{10+20}{2} = 15$ and so on.

18.8 CUMULATIVE FREQUENCY AND CUMULATIVE FREQUENCY TABLE

The cumulative frequency of a class interval is the sum of frequencies of all classes up to that class (including the frequency of that particular class).

Thus cumulative frequency table for example 2 is as follows :

Marks	Frequency	Cumulative Frequency
10 – 20	4	4
20 – 30	9	4 + 9 = 13
30 – 40	5	4 + 9 + 5 = 18
40 – 50	6	4 + 9 + 5 + 6 = 24

This table can also be written in the form of **'less than'** cumulative frequency table as :

Marks	Cumulative Frequency (*i.e.* No. of students)
less than 10	0
less than 20	4
less than 30	13
less than 40	18
less than 50	24

EXERCISE 18(A)

1. State, which of the following variables are continuous and which are discrete :
   (a) number of children in your class.
   (b) distance travelled by a car.
   (c) sizes of shoes. (d) time.
   (e) number of patients in a hospital.

2. Given below are the marks obtained by 30 students in an examination :

08	17	33	41	47	23	20	34
09	18	42	14	30	19	29	11
36	48	40	24	22	02	16	21
15	32	47	44	33	01		

   Taking class intervals 1– 10, 11 – 20, ..., 41 – 50; make a frequency table for the above distribution.

3. The marks of 24 candidates in the subject mathematics are given below :

45	48	15	23	30	35	40	11
29	0	3	12	48	50	18	30
15	30	11	42	23	2	3	44

   The maximum marks are 50. Make a frequency distribution taking class intervals 0 – 10, 10 – 20, ...

4. Fill in the blanks :
   (a) A quantity which can vary from one individual to another is called a ....
   (b) Sizes of shoes are .... variables.
   (c) Daily temperature is .... variable.
   (d) The range of the data 7, 13, 6, 25, 18, 20, 16 is ....

Range of the data is the difference between the highest and the lowest values.

(e) In the class interval 35 – 46; the lower limit is .... and upper limit is ....

(f) The class mark of class interval 22 – 29 is ....

5. Find the actual lower class limits, upper class limits and the mid-values of the classes : 10 – 19, 20 – 29, 30 – 39 and 40 – 49.

6. Find the actual lower and upper class limits and also the class marks of the classes : 1·1 – 2·0, 2·1 – 3·0 and 3·1 – 4·0.

7. Use the table given below to find :

(a) The actual class limits of the fourth class.

(b) The class boundaries of the sixth class.

(c) The class mark of the third class.

(d) The upper and lower limits of the fifth class.

(e) The size of the third class.

Class Interval	Frequency
30 – 34	7
35 – 39	10
40 – 44	12
45 – 49	13
50 – 54	8
55 – 59	4

8. Construct a cumulative frequency distribution table from the frequency table given below :

(i)

Class Interval	Frequency
0 – 8	9
8 – 16	13
16 – 24	12
24 – 32	7
32 – 40	15

(ii)

Class Interval	Frequency
1 – 10	12
11 – 20	18
21 – 30	23
31 – 40	15
41 – 50	10

9. Construct a frequency distribution table from the following cumulative frequency distribution:

(i)

Class Interval	Cumulative Frequency
10 – 19	8
20 – 29	19
30 – 39	23
40 – 49	30

(ii)

C. I.	C. F.
5 – 10	18
10 – 15	30
15 – 20	46
20 – 25	73
25 – 30	90

10. Construct a frequency table from the following data :

Marks	No. of students
less than 10	6
less than 20	15
less than 30	30
less than 40	39
less than 50	53
less than 60	70

11. Construct the frequency distribution table from the following cumulative frequency table:

Ages	No. of students
Below 4	0
Below 7	85
Below 10	140
Below 13	243
Below 16	300

(i) State the number of students in the age group 10 – 13.

(ii) State the age-group which has the least number of students.

12. Fill in the blanks in the following table:

Class Interval	Frequency	Cumulative Frequency
25 – 34	....	15
35 – 44	....	28
45 – 54	21	....
55 – 64	16	....
65 – 74	....	73
75 – 84	12	....

13. The value of π upto 50 decimal places is :
3·14159265358979323846264338327950288419716939937510

(i) Make a frequency distribution table of the digits from 0 to 9 after the decimal place.

(ii) Which are the most and the least occuring digits ?

18.9 GRAPHICAL REPRESENTATION OF DATA

While reading a daily newspaper or a scientific article, most people either skip over the column of figures or at the most get only a feeble idea about them. But when these figures are graphically (pictorially) represented, it becomes more noticeable and eye-catching, leaving a more lasting effect on the mind of the reader. Of course, a graphical representation should be properly titled and labelled so as to convey to the reader, what it is about.

There are various methods of representing the numerical data graphically. Some of these are given below :

18.10 GRAPHICAL REPRESENTATION OF CONTINUOUS FREQUENCY DISTRIBUTION

1. Histogram :

In this case, the rectangles are drawn with class-intervals as bases and their heights are proportional to the frequencies of respective classes.

Steps :

1. Convert the data into the exclusive form, if it is in the inclusive form.
2. Taking suitable scales, mark the class-intervals on x-axis and frequencies on y-axis.

The scales chosen for both the axes need not be the same.

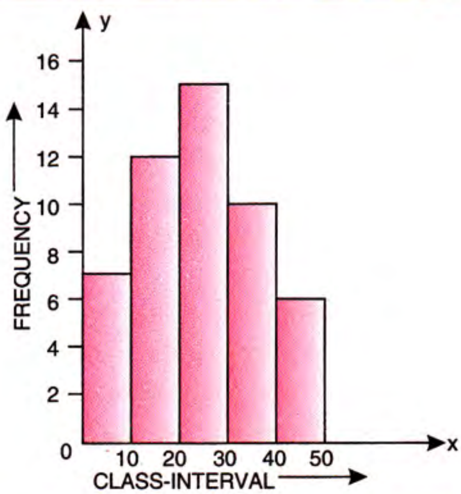
3. Construct rectangles with class-intervals as bases and the corresponding frequencies as heights.

3 Draw a histogram to represent the following :

Class-interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency	7	12	15	10	6

Solution :

In a bar chart, the height of each bar matters and not its width. Whereas, in histogram, the height as well as width of each rectangle matter.

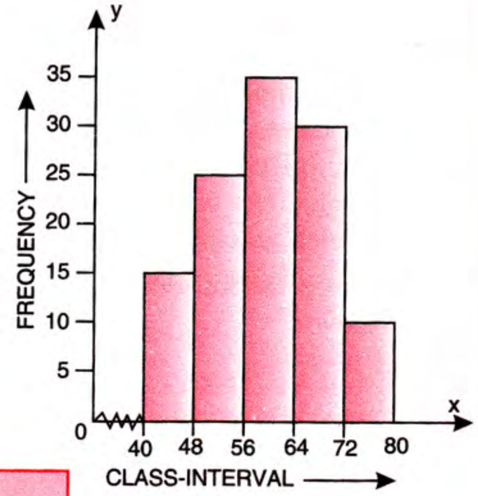


4 Draw a histogram to represent the following :

Class-interval	40 – 48	48 – 56	56 – 64	64 – 72	72 – 80
Frequency	15	25	35	30	10

Solution :

> Since the scale on x-axis starts at 40; a kink (break) or a zig-zag curve is shown near the origin to indicate that the graph is drawn to scale beginning at 40 and not at the origin itself.



5 Draw a histogram for the following data :

Class-interval	1 – 10	11 – 20	21 – 30	31 – 40
Frequency	7	12	15	13

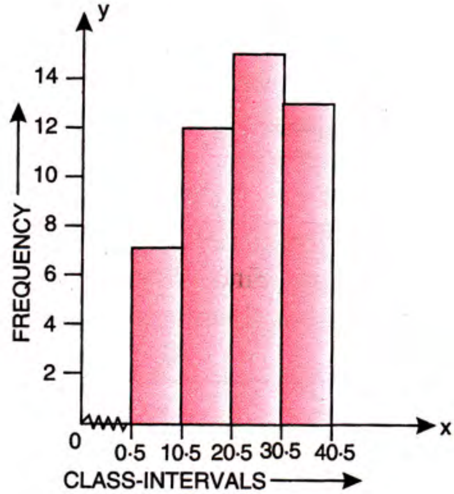
Solution :

In this example, the class-intervals are inclusive. So, first convert the classes into the exclusive form.

On making the classes exclusive (*i.e.* continuous), we get actual class-limits as :

0·5 – 10·5, 10·5 – 20·5, 20·5 – 30·5, 30·5 – 40·5.

Class-Interval	Frequency
0·5 – 10·5	7
10·5 – 20·5	12
20·5 – 30·5	15
30·5 – 40·5	13



2. Frequency Polygon :

When the mid-points of the adjacent class-intervals, of the given frequency distribution, joined by straight line segments, the figure so obtained is called a *frequency polygon.*

To complete the polygon, the mid-points at each end are joined to the immediately lower or higher mid-point at zero frequency, *i.e.* on the x-axis.

A frequency polygon can be drawn : (i) using histogram (ii) without using histogram.

Type 1 : *Using histogram*

Step :

1. Using the given frequency distribution, draw a histogram.
2. Mark the mid-point of upper horizontal line of each rectangle.

3. Mark the mid-point of the class-interval, adjacent to first class-interval, at zero frequency and also mark the mid-point of the class-interval, adjacent to last class-interval, at zero frequency.
4. Join all the mid-points marked to get the required frequency polygon.

6 Draw a frequency polygon from the following data, giving the age of doctors working in C.G.H.S. in a city.

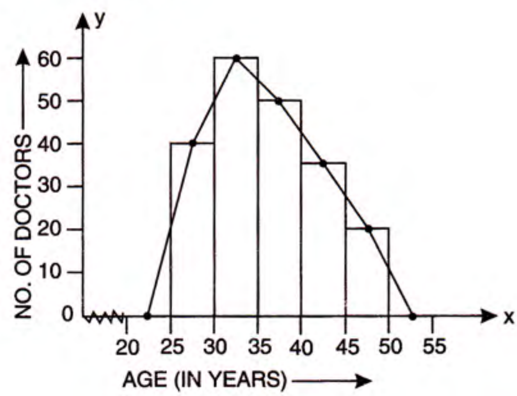
Age (in years)	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50
No. of doctors	40	60	50	35	20

Solution :

The required frequency polygon is shown in the following figure.

Steps :

1. Draw a histogram for the given data.
2. Mark the mid-point at the top of each rectangle of the histogram drawn.
3. Also, mark the mid-point of the immediately lower class-interval (in the given example, the immediately lower class-interval is 20-25) and mid-point of the immediately higher class-interval (in the given example the immediate upper class-interval is 50-55).



4. Join the consecutive mid-points marked by straight lines to obtain the required frequency polygon.

Type 2 : *Without using histogram*

Step :

1. Find the class-mark (mid-value) of each given class-interval.

$$\text{Class-mark} = \text{mid-value} = \frac{\text{Upper limit} + \text{Lower limit}}{2}.$$

2. On a graph paper, mark class-marks along x-axis and frequencies along y-axis.
3. On this graph paper, mark points taking values of class-marks along x-axis and the values of their corresponding frequencies along y-axis.
4. Draw line segments joining the consecutive points marked in step (3) above.

Do not forget to join the mid-value of the class-interval just before the first class and the mid-value of the class-interval just after the last class. This completes the required **frequency polygon**.

7 Draw a frequency polygon from the following frequency distribtuion :

C.I.	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
f	4	8	12	10	7	4

Solution :

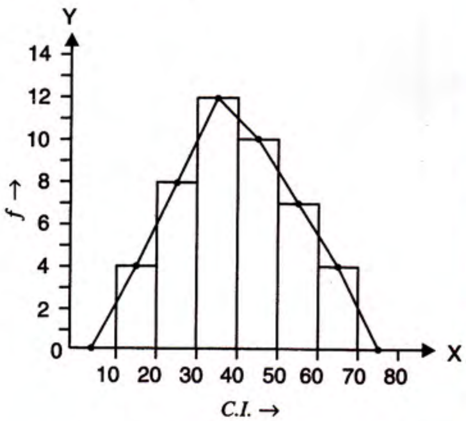
C.I.	f
10 – 20	4
20 – 30	8
30 – 40	12
40 – 50	10
50 – 60	7
60 – 70	4

⇒

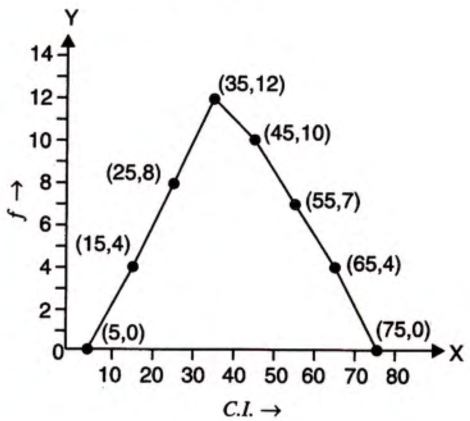
C.I.	Class-mark	f
0 – 10	5	0
10 – 20	$\frac{1}{2}(10 + 20) = 15$	4
20 – 30	25	8
30 – 40	35	12
40 – 50	45	10
50 – 60	55	7
60 – 70	65	4
70 – 80	75	0

∴ The required frequency polygon will be :

Using histogram



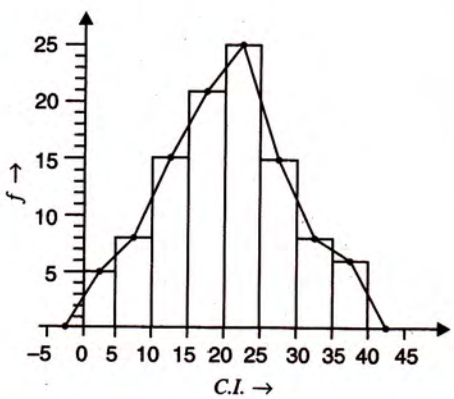
Without using histogram



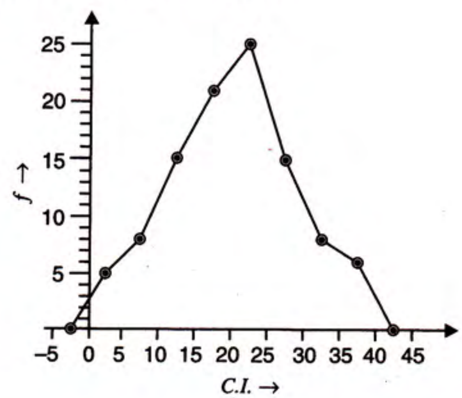
Make the following frequency polygons clear :

C.I.	0– 5	5– 10	10– 15	15–20	20–25	25–30	30–35	35–40
f	5	8	15	21	25	15	8	6

Using histogram



Without using histogram



EXERCISE 18(B)

1. Construct a frequency polygon for the following distribution :

Class-intervals	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24
Frequency	4	7	10	15	11	6

2. Construct a combined histogram and frequency polygon for the following frequency distribution :

Class-intervals	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	3	5	6	4	2

3. Construct a frequency polygon for the following data :

Class-intervals	10 – 14	15 – 19	20 – 24	25 – 29	30 – 34
Frequency	5	8	12	9	4

4. The daily wages in a factory are distributed as follows :

Daily wages (in ₹)	125 – 175	175 – 225	225 – 275	275 – 325	325 – 375
Number of workers	4	20	22	10	6

Draw a frequency polygon for this distribution.

5. Draw frequency polygons for each of the following frequency distribution :

(a) using histogram (b) without using histogram.

(i)

C.I.	10–30	30–50	50–70	70–90	90–110	110–130	130–150
f	4	7	5	9	5	6	4

(ii)

C.I.	5–15	15–25	25–35	35–45	45–55	55–65
f	8	16	18	14	8	2