CHAPTER I

INTRODUCTION AND INTRODUCTORY CONCEPTS

Statistics is a numerical description of some events or subjects. Secondly, it is a method of analysis and interpretation of data. However, statistics usually refers to techniques and methods. That is, statistics is a branch of knowledge which includes appropriate method of collection of data on certain problem, its presentation and analysis, and finding out the truth from the results of the analysis. Universal and complete definition of statistics is not available. Users have defined statistics from their own application view point. Definitions given by three famous scientists are stated below :

According to **Prof. A.L. Bowley**, "Statistics is a science of measurement of social organism regarded as a whole in all its manifestation".

Webster defined statistics as "the classified facts relating to the condition of the people in a state especially those facts which can be stated in numbers or in table of numbers or in any tabular or classified arrangement".

Webster's definition is limited only in the information on the general conditions of the people of a state; it does not include information on the other branches of knowledge. But modern statistics has included all spheres of human activities. According to Bowley's definition, statistical science refers to mutually distributed numerical data in any field of investigation. This definition does not cover the causes that affect the correctness of the information in the data collection process.

However, **Prof. H. Secrist** has given a definition of statistics covering all its application fields. According to him "statistics means the aggregate of facts affected to a considerable extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable levels of accuracy, collected in a systematic method for a predetermined purpose and placed in relation to each other". This can be considered as a complete definition of statistics.

The different functions of statistics are -

- Collection of data
- Organization of data
- Presentation of data
- Analysis of data
- Interpretation and drawing inference.

Thus "**statistics can also be defined as a branch of science which deals with collection, organization, presentation and analysis of numerical data and drawing valid inference therefrom**".

Characteristic Features of Statistics :

In order to help the reader in understanding the nature of the subject some of the characteristic features of the science of statistics are stated below :

i. Statistics deals with aggregate of individuals rather than with individuals. Per capita income of a country is a statistical information because it is an information about the population.

ii. Statistics deals with variation.

iii. Statistics deals with only numerically specified populations.

iv. Statistical inferences are drawn with the probability of uncertainity.

v. The logic used in statistical inference is inductive.

1.1. Uses of Statistics :

Statistics is such a science, application of which is inevitable in all spheres of life. Statistics has wide application in solution of problems related to Economics, Social Science, Biological Science, Agricultural Science, Business, Planning, Education and Research. There is no branch of knowledge whose data analysis does not require statistical techniques. Application of statistics in some important fields are briefly discussed in this section.

(i) Statistics in Agriculture :

Statistics is widely used in the field of agriculture. In any country, particularly in an agro-based country, development planning and its implementation is very important. Agricultural census are conducted to collect different information on agriculture. Data on total cultivable land, cultivated and cultivable crops, fertilizer use, irrigation etc. are collected, organized and presented. Statistical techniques are applied in analysing these data and on the basis of such analysis regionwise requirements of fertilizer, irrigation etc. for different crops are estimated. Besides, statistical techniques are essential for collection and analysis of data on domestic animals and birds, in respect of their numbers, food, health, growth etc. Various information are collected and analysed even on wild animals using statistical techniques and analysis of various data on veterinary treatment, fish culture, etc. Above all, statistics plays a very important role in the overall agricultural development.

(ii) Statistics in Economics :

Statistics has an intensive relationship with Economics. Statistical techniques are widely used in formulation of economic rules and in testing their effectiveness. In fact, statistical methods has made economics more intensified and popularised. Through application of statistical techniques in the field of Mathematical Economics, a new and popular branch of Economics, named Ecomometrics has been created.

(iii) Statistics in Planning :

Developing countries essentially designs development plans for overall national development. The present age, is often termed as the age of planning. Use of statistics is inevitable in designing such development programmes. National policies are prepared through collection and analysis of various data on peoples' standard of living of country, poverty level, education, trade, different management etc. using statistical methods.

(iv) Statistics in Biology :

Statistical techniques are widely used in analysis of various information in different areas of biological sciences; these data relates to births, deaths, breeding, genetics etc. A new branch of statistics named Biometry has been developed to deal with the different biological aspects. Population studies including fertility, mortality, migration etc. involve lot of statistical exercises.

(v) Statistics in Trade and Commerce :

In the present world of competition data collection and analysis are inevitable for progress and development in trade and commerce. Statistical methods are used as essential tools for collection, analysis and interpretation of data on demand & supply of different commodities, trade cycle, consumers taste, principle and purchasing power etc.

1.2. Limitations of Statistics :

In spite of popular uses of statistical methods in different areas of knowledge, there are some limitations of statistics too; those are briefly discussed below :

- Statistics does not refer to the characteristics of any individual, rather it analyses the collected data and refers to the overall results.

- Statistical results are true on the average; for particular case it may not be true. For example suppose in a certain locality 5% crop lands suffers from pest attacks. The crop of a particular land may not be at all infested whereas infestation of crop in some other plots may be 15% or even more.
- Statistics usually collects data through sample survey and comments on the population characteristics on the basis of sample information. Such sample based comments may not be true in some cases.

1.3. Abuses of Statistics :

- Users of statistics must have sufficient expertise on the subject, because improper use of statistical methods may create complicated problems. Untrained and inexperienced users are very likely to take faulty decision.
- Collected data should be analysed through appropriate statistical technique to arrive at correct decision. But favourable decisions are often taken due to ignorance or by purposively manipulation of data.
- Comments made on the basis of insufficient or inadequate data may be faulty.

Therefore, to ensure proper use of statistical methods, the data should be relevant and complete. The users of statistics should also have necessary knowledge and experience on statistical methods.

1.4. Population and Sample :

It is important to distinguish between a population and a sample.

A population (or universe) is defined as the aggregate of the elementary units. Since the number of elementary units is equal to the number of observations, we may say that a population is the aggregate of the observations. For example, assume that 650 students are admitted to the first semester of the first level at BAU. We are interested on their H.S.C. score. Each student is an elementary unit and the 650 students together comprise the population. Or, we may say that 650 observations (H.S.C. scores of 650 students) is the population. The size of the population is usually denoted by N.

A sample is a set of n observations (elementary units) drawn from the population. This n is known as the size of the sample. As an illustration, if we select 25 students from the population of 650, we have a sample of size 25. Or, we may say that we select 25 observations (H.S.C. scores) from the population of 650 scores (sample and sampling methods will be discussed in chapter IX). Unknown characteristics which refer to populations are called parameters; and that related to samples are called statistic. Populations, as the term is used in statistics, are arbitrarily defined groups. They need not be as large as the one used here as illustration.

1.5. Scales of Measurement :

The theory of measurement consists of a set of separate or distinct theories, each concerning a distinct scales of measurement. The operations admissible on a given set of scores depend on the scale of measurement used. Different scales of measurement are :

1. Nominal or classificatory scale
2. Ordinal or ranking scale
3. Interval scale
4. Ratio scale

Nominal or Classificatory Scale :

Nominal scales are used as measures of identity. Numbers may serve as labels to identify items or classes. When numbers or certain symbols are used to identify the groups to which objects under observation belong, these numbers or symbols constitute the nominal or classificatory scale.

In its simplest form, the numbers carried on the back of athletes represent a nominal scale. Other examples of such scales are the classification of individual into categories. For example, a sample of people under study may be sorted in different categories on the basis of religious belief; (i) Muslim; (ii) Hindu; (iii) Christian; (iv) Buddhists etc. Or, they may be classified on the basis of sex, political party membership, rural-urban and the like. Simple statistics are used with nominal data. For example, the number, proportion or percentage. These classes or categories are mutually exclusive. The only relation involved is that of equivalence. That is, the member of any class or category are equivalent in the property being scaled. The relation is symbolized by the sign "=". Under certain conditions, we can test hypothesis regarding the distribution of cases among categories using the nonparametric statistical test χ^2; the most common measure of association for nominal data is the contingency coefficient, C.

Ordinal or Ranking Scale :

When an ordinal scale is used in measurement, numbers reflect the rank order of the individuals or objects. Objects in one category are not just different from those in the other category of the scale, but there exists some kind of relationship to them. Examples of such relations are higher, more preferred, taller, brighter, smaller, harder, softer etc. Such relations may be designated by the symbol ">" or sometimes by "<". The fundamental difference between a nominal scale and an ordinal scale is that the ordinal scale incorporates not only the relation of equivalence "=" but also the relation greater than ">" or smaller than "<".

Ordinal measures reveal, for instance, which person or object is taller or heavier than the other. But measures do not tell how much taller or how much heavier one is from the other. Statistically not much can be done with ordinal measures except to determine the median and partition values and to compute rank correlation coefficients.

Interval Scale :

The interval scale provides numbers that reflect differences among items. The interval scale has all the characteristics of the ordinal scale, and in addition, provides the distance between any two numbers. That is, we know how large are the intervals (distances) between all objects on the scale. An interval scale is characterised by a common and constant unit of measurement which assigns a real number to all pairs of objects in the ordered set.

Examples on such scales are the centigrade and Fahrenheit thermometers, scores on intelligence tests, etc.

In constructing an interval scale, one must not only be able to specify equivalences, as in a nominal scale, and greater than (or less than) relation, as in ordinal scale, but also be able to specify the ratio of any two intervals. The interval scale is a true quantitative scale. Many statistics are used with interval scales : arithmetic mean, standard deviation and the Pearson's product moment correlation coefficient. Parametric statistical tests such as t -test and F-test are applicable.

Ratio Scale :

The basic difference between this and the preceding type is that ratio scales have an absolute zero. It is true that interval scales (e.g.

Fahrenheit and Centigrade) also have zero points, but such points are arbitrarily chosen. Common ratio scales are measures of length, width, weight, capacity, loudness, and so on. When a ratio scale is used numbers reflect ratios among items, and data obtained with such scales may be subjected to the highest type of statistical treatments.

When the data are in terms of meters or centimeters we can say that one length or height is twice or half of that of another. When our measurements are on an interval scale, we cannot do this. For example suppose the maximum temperature today is 60^0; the same day last year it was 30^0. In this case we cannot state that it is twice as warm today as it was on the same date last year. What is the difference between the two conditions ? When where dealing with meter or centimeter, we were using a measuring scale that was based on an absolute zero; in the second case we were using a scale which started 32 degrees below the freezing point of water. When measurements are on the ratio, such meaningful comparisons can be made. As a matter of fact, when data are of this type, all of the mathematical and statistical operations may be made.

Nominal and ordinal measurements are the most common types achieved in behavioral sciences. Data measured by either nominal or ordinal scales should be analysed by the nonparametric methods. Data measured in interval or ratio scales may be analysed by parametric methods, if the assumptions of the parametric statistical model are tenable.

The information in above discussion of different measurement scales and kinds of statistics and statistical tests appropriate to each scale (when the assumptions of the relevant test are met) are summarised below :

Table 1.1 : Scales of Measurement, Appropriate Statistics and Appropriate Statistical Tests

Scale	Defining relation	Examples of appropriate statistics	Appropriate statistical tests
Nominal	(i) Equivalence	Mode Frequency Contingency coefficient	Nonparametric statistical tests
Ordinal	(i) Equivalence (ii) Greater than or less than	Median Quartiles Rank correlation coeff. Kendall's ω	-Do-
Interval	(i) Equivalence (ii) Greater/less than (iii) Known ratio of any two intervals	Mean Standard deviation Simple and multiple correlation coefficients	Nonparametric and parametric statistical tests
Ratio	(i) Equivalence (ii) Greater/less than (iii) Known ratio of any two intervals (iv) Known ratio of any two scale values	Geometric mean Coefficient of variation	-Do-

1.6. Some Statistical Symbols :

We have already mentioned that the unknown characteristics of a population are called parameters, whereas the characteristics of a sample are called statistic. Some usual symbols for parameters and statistics are -

Table 1.2. Population and Sample Characteristics and Their Notations.

Characteristic	Parameter	Statistic
Mean	μ	$\overline{x}$
Standard deviation	σ	s
Variance	σ^2	s^2
Correlation Coefficient	ρ	r
Regression coefficient	β	b

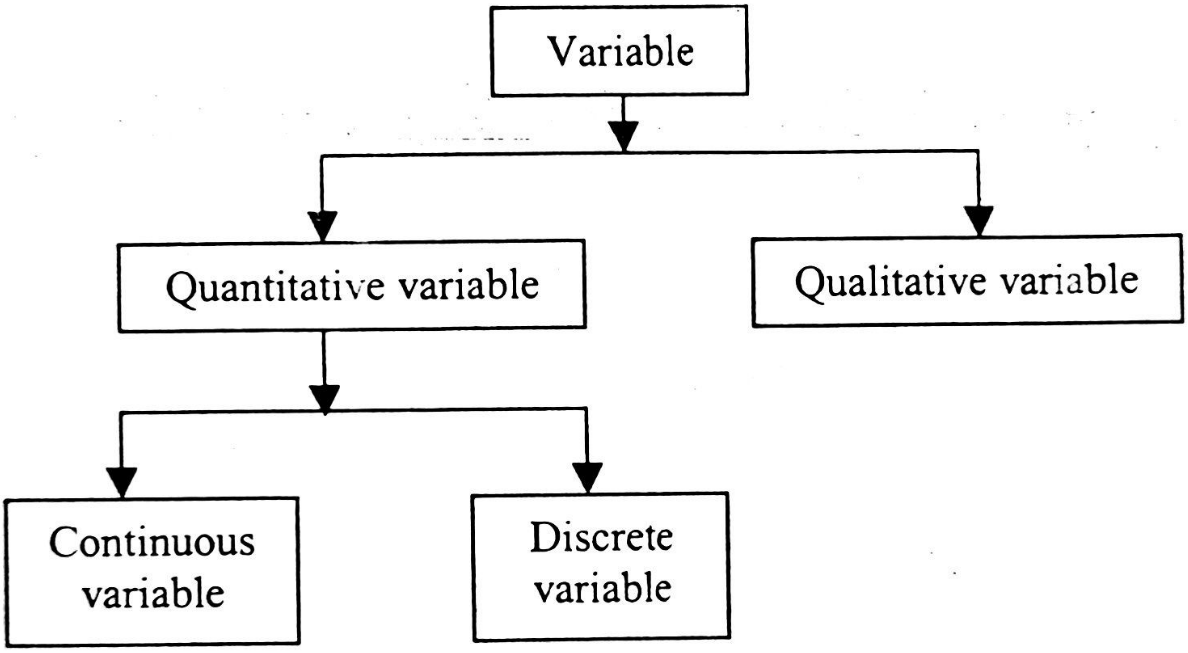
CHAPTER II

VARIABLE AND FREQUENCY DISTRIBUTION

2.1. Variable :

Measurable characteristics of a population that may vary from element to element either in magnitude or in quality are called *variables*. Variables are of two types - *quantitative variable* and *qualitative variable*.

Variables and its classification can be demonstrated as shown below :



Quantitative Variable :

Variable characteristics, whose values are expressed numerically, are known as quantitative variables: Height or weight of students, length or breadth of fishes, weight of tomato, number of grapes per bunch, number of grains per panicle, etc. are some examples of quantitative variables.

Qualitative Variable :

Some variables, which express the quality of population elements, cannot be numerical measured but can be classified or categorised, these are called qualitative variables. For example, merit of students, educational attainment, type of farmers (big, medium, small), type of fishes (sea fish, river fish) etc. cannot be numerically measured but can be grouped into classes or categories. Qualitative variables are also known as attributes.

Quantitative variables are of two types - continuous and discrete.

A variable which can assume any value, integral or fractional, within specified limits, is called a *continuous variable*. For example, height of students, weight of tomato, length of fish, height of trees, weight of animal etc. are continuous variables which can take both integral or fractional values.

On the other hand some variables can take only integral values e.g., number of grains per panicle, number of students per class, number of fishes caught per unit time etc. These are called *discrete variables*.

2.2. Frequency and Frequency Distribution :

In many situations several population elements assume same values; i.e., numerical values of population characteristics may often repeated again and again. For example, several panicles may contain the same number of grains, a number of fishes may have the same length or weight, several students in a class may have the same height. Such repeatation of the value of variable is called frequency. That is, the number of times a particular observation occurs in the data set is the frequency of that particular observation. Frequency is usually denoted by 'f'.

Frequency Distribution :

Information collected in any process are usually classified or grouped according to specific characteristics. Arrangement of observational data according to frequencies of the observations is called *frequency distribution*.

Frequency distribution should be such that the arrangement according to the observations becomes easily understandable. Frequency distributions are constructed mainly to present the data in condensed form and for easy understanding. Frequency distribution is very important in statistical studies.

Construction of Frequency Distributions :

Steps in constructing a frequency distribution are discussed below :

1. Finding the Range : In constructing frequency distribution the highest and the lowest value in the data set are first identified and their difference is obtained. This difference between the highest value and the lowest value is called the *range* usually denoted by R.

Range = Highest value - Lowest value.

2. Decision About the Number of Classes : After finding the range, it is necessary to decide the number of classes in which the entire data set should be divided. Choice of the number of classes should be realistic; this number should not be very small and at the same time it should not be very large so that the aim of construction of frequency distribution (condensation) is not achieved. It is generally expected to limit the number of classes between 7 and 15. There is no hard and fast rule for choosing the number of classes. However, M.A. Sturge's formula gives a guideline for desired number of classes. The formula is -

$$k = 1 + 3.322 \log_{10} N$$

where N is the total number of observations in the data set and k is the desired number of classes.

3. Choosing the Class Interval : The next step of constructing frequency distribution is the calculation of the class interval. Each class will have two limits, the lower limit (the lower value) and the upper limit (the higher value). The difference of the upper limit and the lower limit of a class is known as *class interval*, usually denoted by c or h. If the range is divided by the number of classes, we get the class interval.

$$\text{Class Interval (C)} = \frac{\text{Range}}{\text{No. of classes}}$$

The value of c is taken as the next integral value of the ratio R/k. For choosing the class interval, there is no rigid rule as to use the exact end values of the data set, rather convenient values near the highest and lowest observations of the data set may be used. However, class interval should be such that classes are distinct and separate from each other. Depending on the nature of the variable, two different methods are used in choosing the class limits. If the variable is discrete, closed intervals like $a \leq x \leq b$ are used, both the lower and upper limits are included (e.g., 0-4, 5-9, 10-14, ..., etc.). On the other hand, if the variable is continuous, open interval system ($a < x \leq b$ or $a \leq x < b$) is used; one of the class limits is included and the other is excluded (usually the lower limit is included). In this case the classes will be 0-5, 5-10, 10-15, ..., etc.

4. Counting of Frequencies : For convenience of counting the number of observations falling within each class tally marks are used; frequency of each class is determined by counting the tally marks.

Sometimes it may be necessary to know the observations greater or smaller than a particular value or class of values. For this, cumulative frequencies for observation or class are obtained.

Example 2.1 :

Suppose the marks obtained by 50 students in an examination in Economics are as follows :

32 27 19 40 31 17 15 18 21 27 38 15 33 34 29
26 16 25 33 36 24 22 26 19 36 18 25 20 25 25
31 24 16 28 30 24 29 42 29 28 26 27 47 43 22
25 28 22 24 23

Here the variable is the marks obtained by the students. The data as shown above are called raw or ungrouped data.

If it is needed to describe the performance of the students, it may be done in a number of ways.

We may enumerate the grade of each student either in ascending or descending order: data such arranged are said to be arranged in array. Counting the number of times each value of the variable occurs, we get a table of the following type :

Table 2.1 Frequency Distribution of Marks

Marks	Frequency (No. of students)	Cumulative frequency	Marks	Frequency (No. of students)	Cumulative frequency
15	2	2	28	3	33
16	2	4	29	3	36
17	1	5	30	1	27
18	2	7	31	2	39
19	2	9	32	1	40
20	1	10	33	2	42
21	1	11	34	1	43
22	3	14	36	2	45
23	2	16	38	1	46
24	4	20	40	1	47
25	5	25	42	1	48
26	3	28	43	1	49
27	2	30	47	1	50

Such a table is known as frequency table or frequency distribution. The above arrangement is an improvement over the raw data, but to get a still better idea of the performance of the students we reclassify the data into grouped frequency distribution as shown below :

Table 2.2: Grouped Frequency Distribution of Marks

Class interval	Tally mark	Frequency (No. of students)	Cumulative frequency	
			Ascending	Descending
15-20	卌 ||||	9	9	50
20-25	卌 卌 |	11	20	41
25-30	卌 卌 卌 |	16	36	30
30-35	卌 ||	7	43	14
35-40	|||	3	46	7
40-45	|||	3	49	4
45-50	|	1	50	1

This type of classification of raw data is called grouped frequency distribution or simply frequency distribution.

In the above example the highest value is 47, the lowest value is 15 and the range is, R = 47-15 = 32.

According to Sturge's formula,

$$k = 1 + 3.322 \log_{10} 50 = 6.47$$

That is, 6 to 7 groups are appropriate in this case.

Again. $C = \frac{R}{k} = \frac{32}{7} = 4.57$; accordingly 5 is taken as the class interval.

Example 2.2: Weight (in gm.) of tomato harvested from the kitchen garden are given below :

75	80	52	87	95	105	92	82	120	65
55	100	115	92	82	97	85	72	67	98
115	62	85	98	110	105	77	63	80	90
54	89	108	103	75	53	105	117	95	64
77	85	94	72	68	100	78	89	94	102
82	95	98	100	77	85	92	97	72	85
72	83	66	58	96	75	88	90	80	95
63	78	84	92	88	77	65	85	92	87

In constructing a frequency distribution the highest and the lowest observations are to be identified first. In the present data set the highest value is 120 and the lowest value is 52. Therefore range is

$$R = 120 - 52 = 68 \; ; \; N = 80$$

According to Sturge's formula

$$k = 1 + 3.322 \log_{10}80 = 1 + 3.322 \times 1.903089987 = 7.322.$$

The next integral value of k is 8; the data set may be grouped in about 8 classes.

$$\text{Now, } C = \frac{R}{K} = \frac{68}{8} = 8.5 \approx 9$$

It will be convenient to take 10 as the class interval. As the variable here is continuous, open interval method is to be followed in grouping the data set. Though the lowest observation is 52, it is convenient to start from 50. The classes or groups will, therefore, be 50-60, 60-70, 70-80, 80-90, 90-100, 100-110 and 110-120.

The frequency distribution will be -

Table 2.3: Frequency Distribution of Weight of Tomato

Class interval	Tally mark	Frequency	Cumulative frequency	
			Ascending	Descending
50-60	||||̸	5	5	80
60-70	||||̸ ||||	9	14	75
70-80	||||̸ ||||̸ |||	13	27	66
80-90	||||̸ ||||̸ ||||̸ ||||̸	20	47	53
90-100	||||̸ ||||̸ ||||̸ ||||	19	66	33
100-110	||||̸ ||||	9	75	14
110-above	||||̸	5	80	5

The highest observation 120 appears only once in the data set. This 120 could make another group 120-130. Without introducing a separate group for only one observation, it has been included in the preceding group making it 110-above, instead of 110-120.

Example 2.3 : The following data show the number of grapes per bunch :

25	75	15	20	18	62	45	33	40	45
77	30	35	25	65	42	55	37	44	50
40	35	38	47	52	45	33	28	22	22
18	29	48	55	60	58	43	40	47	39
35	45	43	52	57	50	48	55	59	42
28	36	54	48	58	68	78	61	53	42

Here, N = 60, the highest value is 78 and the lowest value is 15. The variable is discrete in nature.

$$R = 78 - 15 = 63 \text{ and } k = 1 + 3.322 \log_{10}60 = 6.91 \approx 7.$$

The data may be classified in more or less 7 groups.

Again, $C = \frac{R}{k} = \frac{63}{7} = 9$

For convenience, however, 10 may be taken as the class interval.

Table 2.4 : Frequency Distribution of No. of Grapes per Bunch

Class interval	Tally mark	Frequency	Cumulative frequency	
			Ascending	Descending
15-24	~~IIII~~ I	6	6	60
25-34	~~IIII~~ III	8	14	54
35-44	~~IIII~~ ~~IIII~~ ~~IIII~~	15	29	46
45-54	~~IIII~~ ~~IIII~~ ~~IIII~~ I	16	45	31
55-64	~~IIII~~ ~~IIII~~	10	55	15
65-74	II	2	57	5
75-84	III	3	60	3

2.3. Graphical Representation of Frequency Distribution :

Frequency distributions may be presented by graphs and charts in order to make them more clear, more easily understandable and to compare distributions quickly. It is also easy to understand by illiterate persons and people from different regions with different languages. Graphical representation brings to light the salient features of the data at a glance. It is also useful in locating some partition values.

The following graphs are generally used in representing frequency distributions :

1. Dot Frequency Diagram
2. Histogram and Bar Diagram

3. Frequency Polygon and Frequency Curve
4. Cumulative Frequency Curve or Ogive
5. Pie Chart

Dot Frequency Diagram:

The X-axis is used for the variable values and the Y-axis is for the frequency; if we indicate the frequencies of each variable value by dots, the resulting diagram is known as dot frequency diagram.

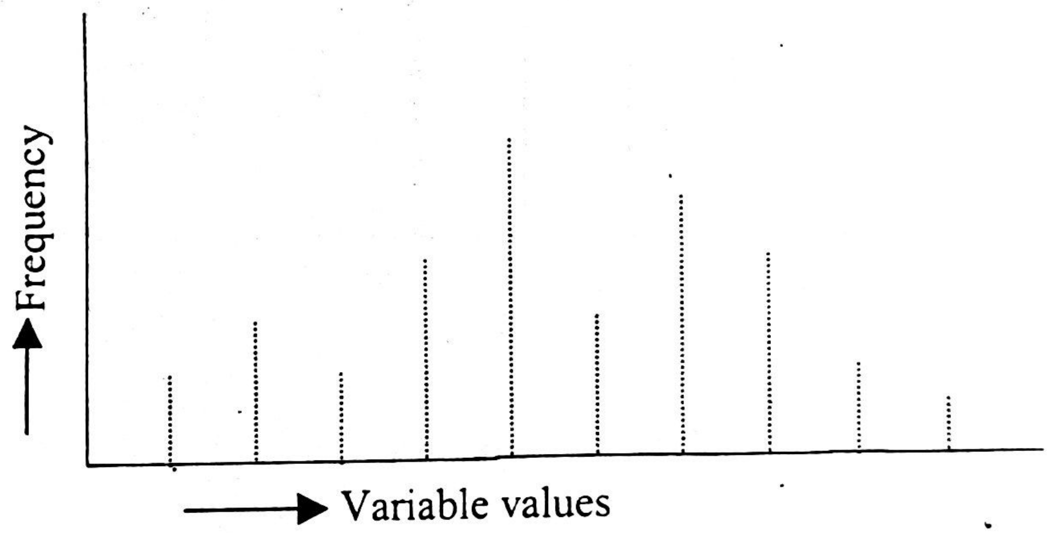


Fig. 2.1: Dot frequency diagram.

Histogram :

Class intervals are plotted along the X-axis and frequencies are plotted along the Y-axis. For each class or group, a rectangle is drawn taking class interval as the base and the class frequency as the height. For continuous variable, the rectangles such drawn are attached to adjacent rectangles at both sides and the resulting graph is known as histogram.

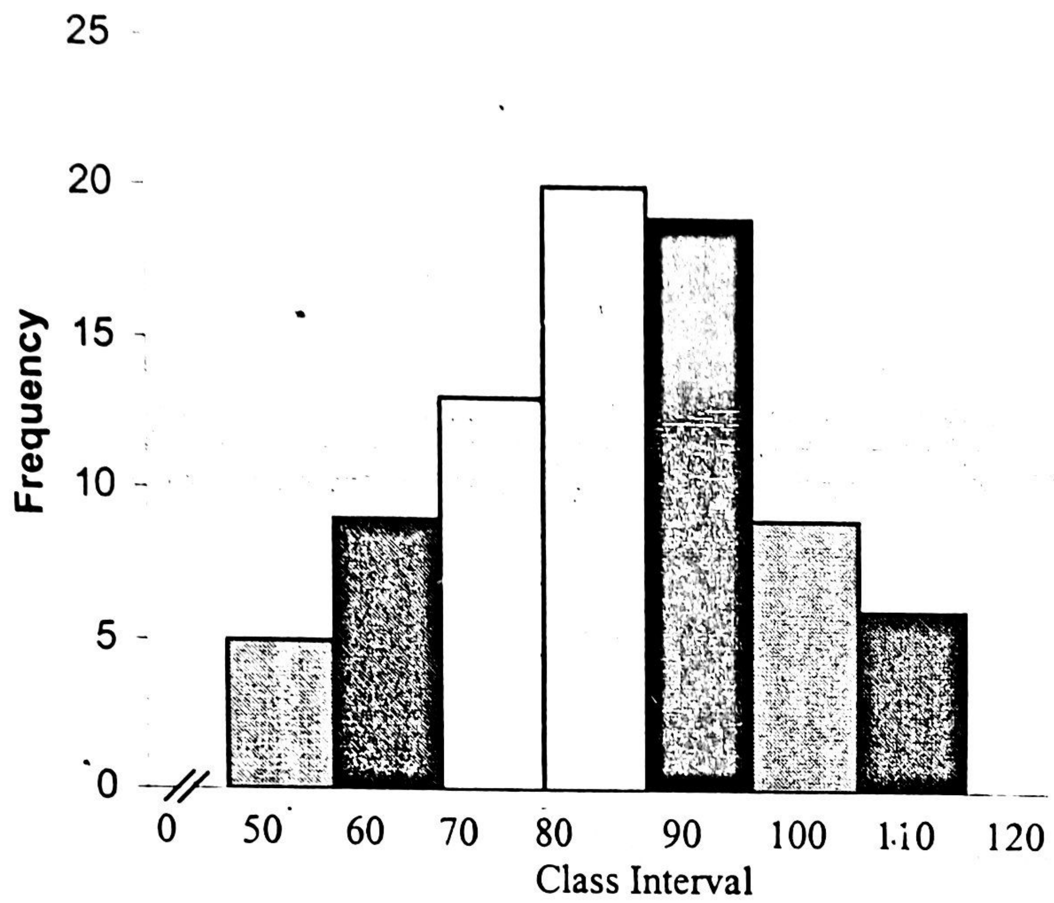


Fig. 2.2: Histogram for data in table 2.3.

For drawing histograms of frequency distributions having unequal class intervals, frequency density, instead of frequency is plotted along the Y-axis.

Frequency density is obtained as $f_d = \frac{f}{c}$; c being the class interval.

Example 2.4:

Drawing the histogram of the distribution of members per family in a certain locality is described below :

Family size (class Interval)	No. of families (f)	Class interval (C)	Frequency density $f_d = \frac{f}{c}$;
0-2	8	2	4
2-4	14	2	7
4-8	16	4	4
8-12	20	4	5
12-20	8	8	1
Total	66		

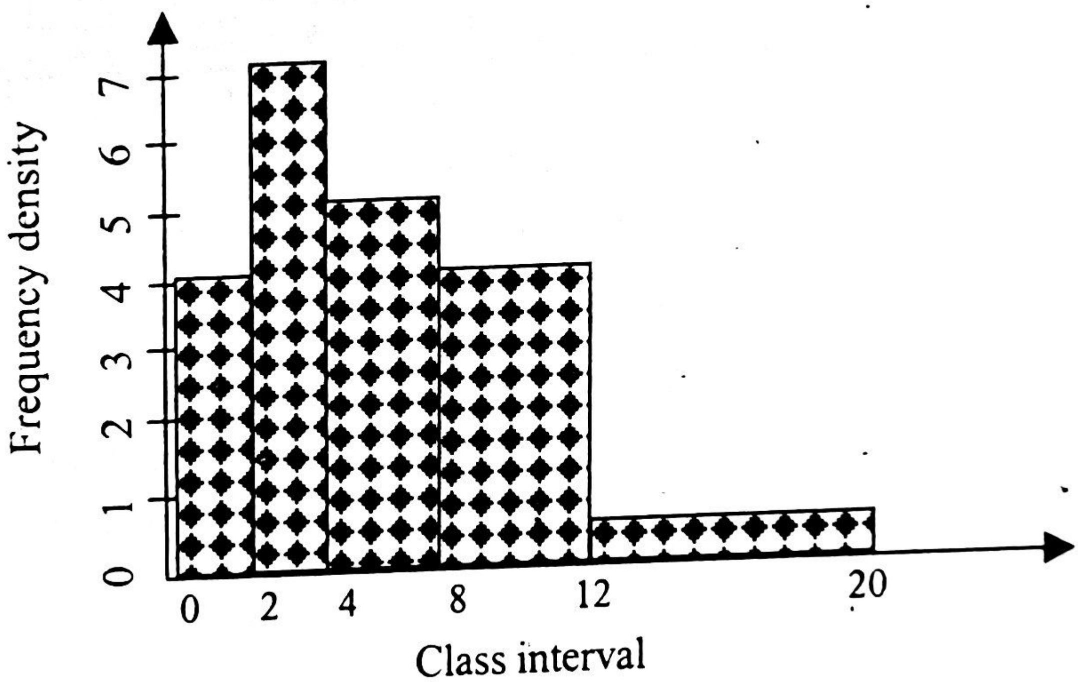


Fig. 2.3: Histogram for data with unequal class interval (Example 2.4)

Bar Diagram : Bar diagram is used mainly to represent discrete frequency distributions. Drawing process of bar diagram is similar to that of histogram. For discrete variables a gap exists between the upper limit of a class and the lower limit of the following class and the adjacent rectangles are not attached to each other. The graph is known as bar diagram.

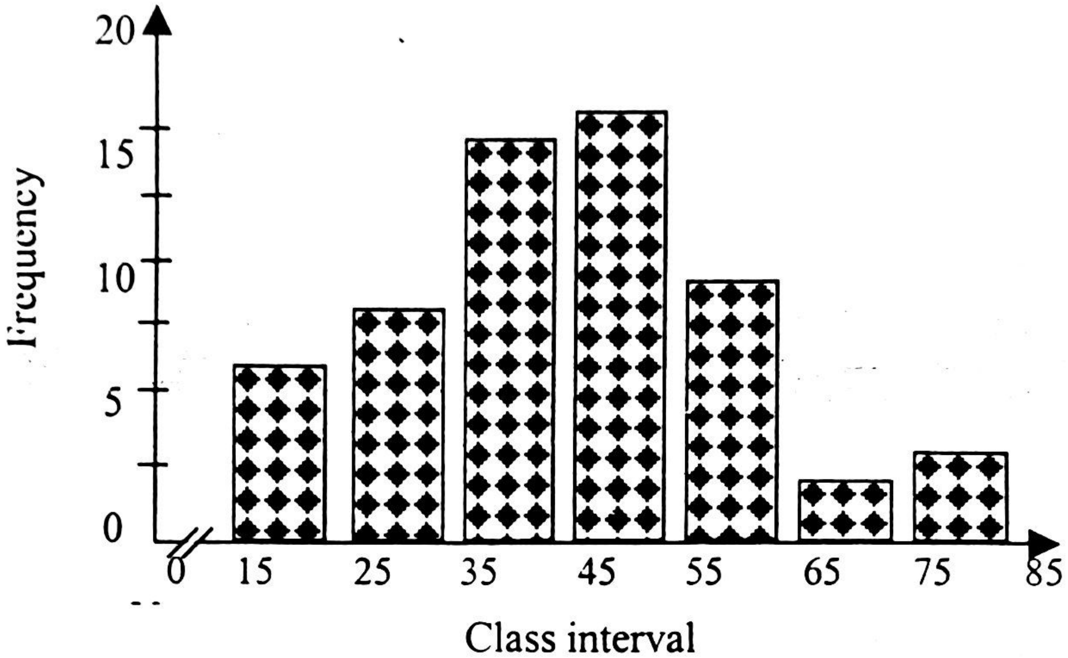


Fig. 2.4: Bar diagram for data in table 2.4.

Besides discrete frequency distributions, bar diagrams can also be used to represent information based on different times or places. For example, crop production in different years, or rainfall in different countries or at different regions of a country may be presented by bar diagrams.

Example 2.5:

Data on annual rainfall at divisional cities of Bangladesh are as follows :

Division	Annual Rainfall (mm.)
Dhaka	1540
Chittagong	2260
Khulna	1159
Rajshahi	1142
Sylhet	3568
Barisal	1200

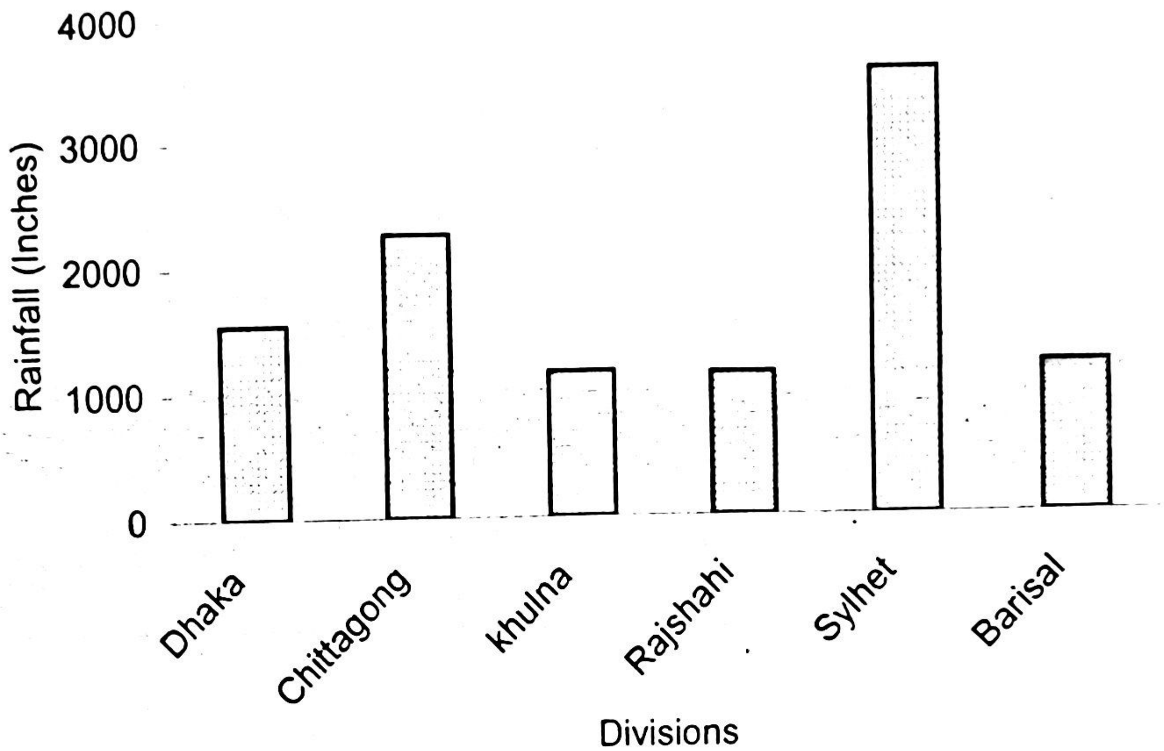


Fig. 2.5: Bar diagram of rainfall at different divisions

Multiple Bar Diagram :

Data on several variables in respect of different places or time-points may be represented by multiple bar diagrams. The simple bars for the variables corresponding to a place or time-point are constructed side by side (without gap). Heights of these bars indicate the values of the respective variables. Usually different bars at the same place or time-point are given with different colours or marks for identification.

Example 2.6:

Data on production of different pulses (in '000 tons) in Bangladesh during the years from 1991-92 to 1994-95 and the corresponding multiple bar diagram are shown below :

Yield of Pulses (000 tons)

Pulses	Year			
	1991-92	1992-93	1993-94	1994-95
Kheshari	185	172	188	189
Moshur	153	163	168	168
Mug & Mashkalai	82	82	82	85

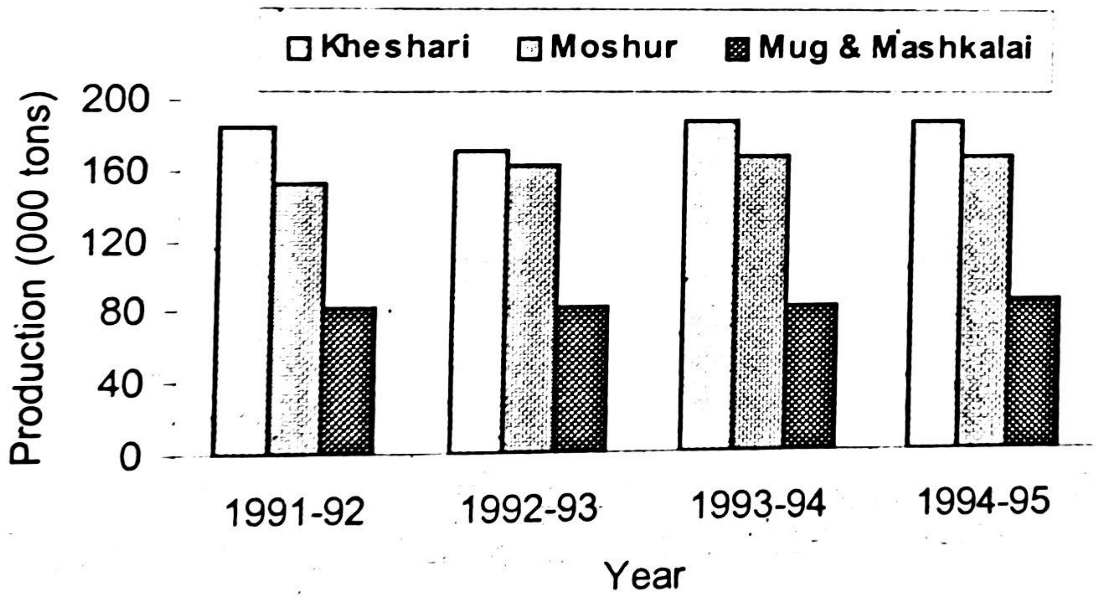


Fig. 2.6: Multiple bar diagram of pulses production.

Comparison between Histogram and Bar diagram

Histogram	Bar diagram
Histograms are used to represent continuous frequency distributions.	Bar diagrams are used to represent discrete frequency distributions.
Histograms are used to represent frequency distribution only.	Besides frequency distributions, data on different places or time points can be represented by bar diagrams.
In drawing the rectangles of the histogram both breath and length of the rectangles are considered.	In drawing bar diagrams consideration of the breath of bars is not necessary. For descent pictorial presentation, bars of suitable breath may be drawn.
Frequency distributions having unequal class intervals may be represented by histograms; in such case the breath of rectangles are unequal.	Bar diagrams are not usually drawn to represent frequency distributions having unequal class intervals.

Frequency Polygon and Frequency Curve :

Mid values of the class intervals are plotted along the X-axis and points are marked on the basis of respective class frequencies plotted along the Y-axis. If the consecutive points are connected with straight line, the resulting graph is a frequency polygon. At the starting and at the end of the distribution one class each are assumed and the polygon, for its completion, is extended upto the mid points of these class intervals, taking their frequencies to be zero.

Frequency polygon for table 2.3 is shown below :

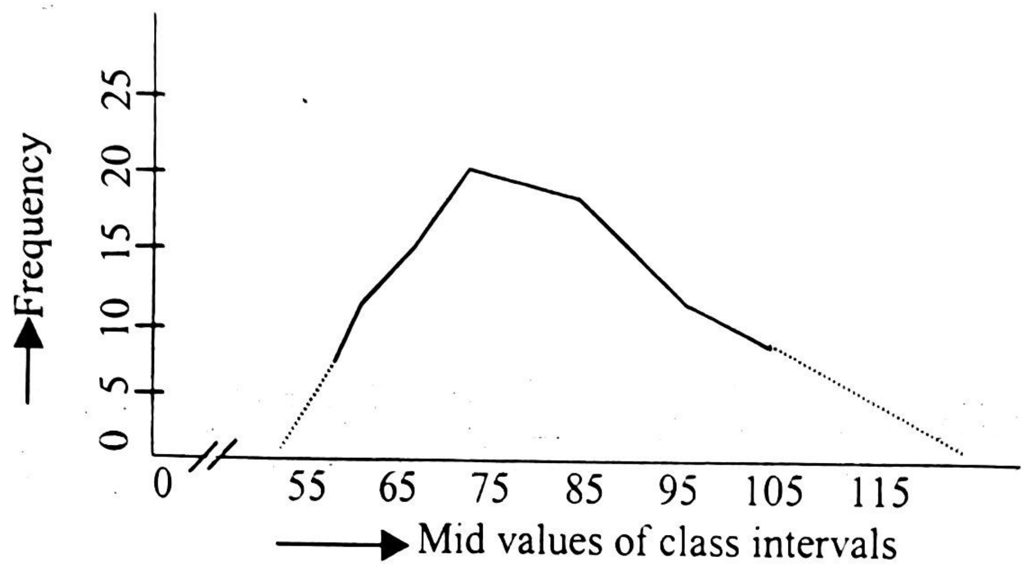


Fig. 2.7: Frequency polygon

Frequency polygon can also be obtained by joining the mid points of the vertical lines of the rectangles of the histogram.

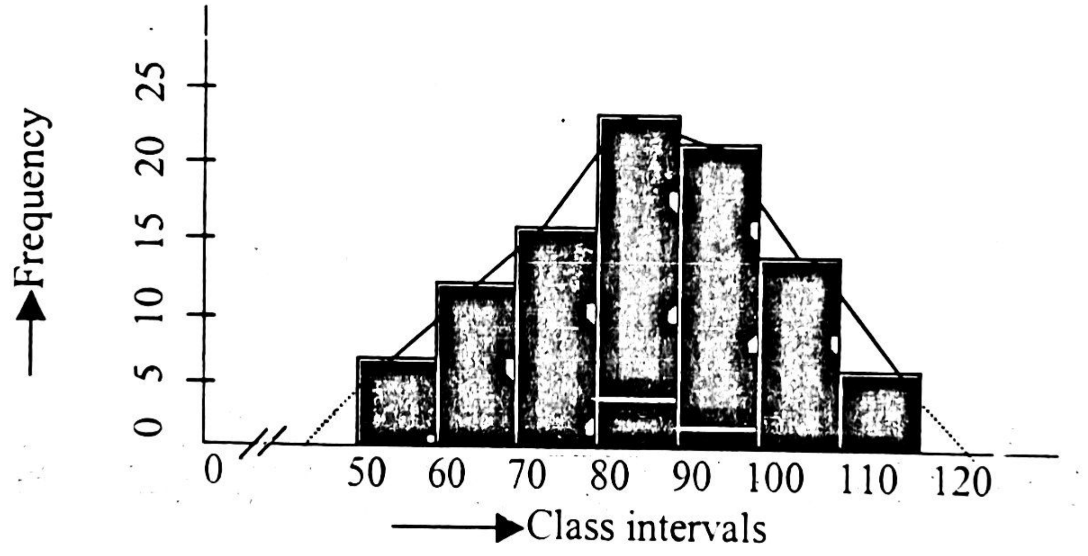


Fig. 2.8: Drawing frequency polygon from histogram in fig. 2.2

Frequency Curve :

In drawing frequency polygons, the consecutive points are connected by straight lines. If the points are connected by a free hand smooth curve, the resulting graph is known as frequency curve. As the frequency curve is a free hand curve, usually it does not pass through all the points; of course, it is desirable to touch the majority number of points. The frequency curve for the table 2.3 is shown below :

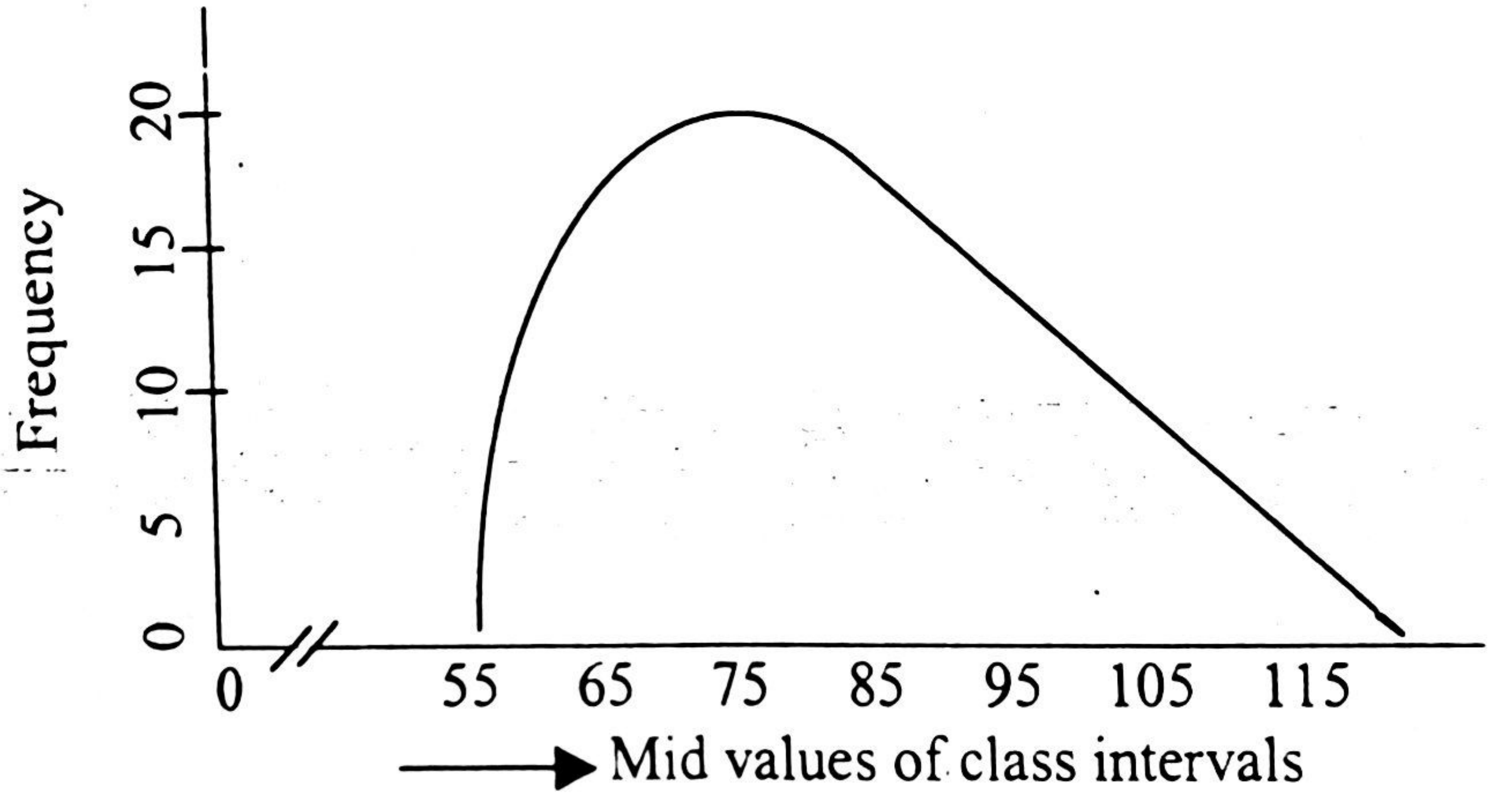


Fig. 2.9: Frequency curve

Pie Chart :

Different components of data are exhibited by splitting a circle. The angle at the centre of a circle is proportionately divided and accordingly splitting the circle we exhibit different components of the data. The division is also done in percentage according to the relative magnitude of different components. Usually different divisions are demarcated by different colours or symboles.

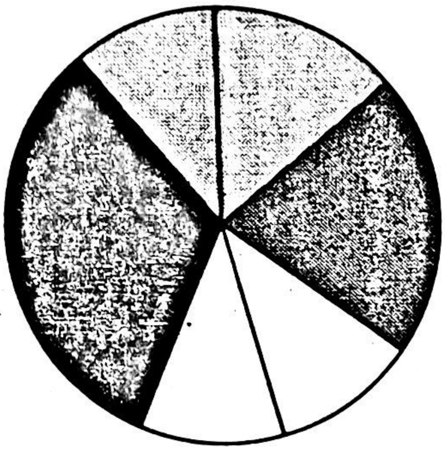


Fig. 2.18 : Pie chart of example 2.5

Exercises

1. The following data show the number of grains per panicle of a certain variety of rice; counts made on randomly selected panicles in an experimental plot :

135	120	125	140	145	105	110	115	118	128
125	130	128	115	110	108	138	140	129	126
102	135	126	129	135	128	126	119	122	130
126	129	132	121	136	139	142	145	148	107
132	127	119	128	130	145	126	103	118	128

a) Present the data in a frequency table taking a suitable class interval.

b) Draw a (i) Bar diagram and
(ii) Frequency polygon