CHAPTER # 2

PRESENTATION OF DATA

2.1 INTRODUCTION

We have discussed the collection of data in the previous chapter. When the data have been collected, it must be presented in a form which is easy to understand. For this purpose following methods are applied.

1. Classification
2. Tabulation
3. Diagrammatic and Graphic representation

2.2. CLASSIFICATION

The process of arranging the data into homogenous classes according to their resemblances and similarities is called classification of data. *e.g.*, the process of sorting letters in a post office. The letters are classified according to cities first then further arranged according to towns, sectors and streets.

2.2.1 The Basis of Classification

The basis of classification are

(i) Geographical or Spatial
(ii) Chronological or Temporal
(iii) Qualitative
(iv) Quantitative

(i) Geographical or Spatial Classification

In geographical classification, the data are classified by geographical regior or locations. For example, population of a country may be classified by provinc division, district and Tehsils etc.

Population size by Province (1998 Census)

Administration Unit	Population (in thousands)
Punjab	72585
Sindh	29991
N.W.F.P.	17555
Balochistan	6511
Fata	3138
Islamabad	799
Pakistan	130580

Sources : Population Census Report, 1998.

(ii) Chronological (Temporal) Classification

When the data are classified according to the time of its occurrence such as year, months, weeks, days, hours etc. It is called chronological classification. An arrangement of data by their time of occurrence is called a time series.

Population of Punjab Province Since 1951

Years	Population
1951	20,556,800
1961	25,499,876
1972	37,611,668
1981	47,292,441
1998	72,585,430

Source : Population Census Report, 1998.

(iii) Qualitative Classification

In qualitative classification, the data are classified on the basis of some quality or attribute, such as religion, sex, beauty, martial status, colour etc.

(iv) Quantitative Classification

In this type of classification data are classified according to some quantitative measurements *e.g.* heights, income, age, sales, temperature etc.

2.3 TABULATION

The systematic arrangement of data in to rows and columns is called tabulation.

2.3.1 The Main Parts of a Table

The main parts of table is

(i) Title
(ii) Prefatory notes
(iii) Column captions and box head
(iv) Row captions and Stub.
(v) Body of a table
(vi) Foot Notes
(vii) Source Notes

(i) Title

A tile is the main heading of the table. Title should be in capital letters and must be written on the top of the table. The title should be clear, brief and self-explanatory.

(ii) Prefatory Notes

It is a statement, given below the title and enclosed in brackets usually describing the units of measurements.

(iii) Column Captions and Box Head

The heading of each column is called a column caption, while the portion of a table that contains the column caption is a called box head. The headings should be clear. Only the first word in each column should be in capital letters.

(iv) Row Caption and Stub.

The heading of each row is called a row caption, while the section of a table that contains row captions is called stub.

(v) Body of a Table

This is the main part of the table, which contains the numerical information.

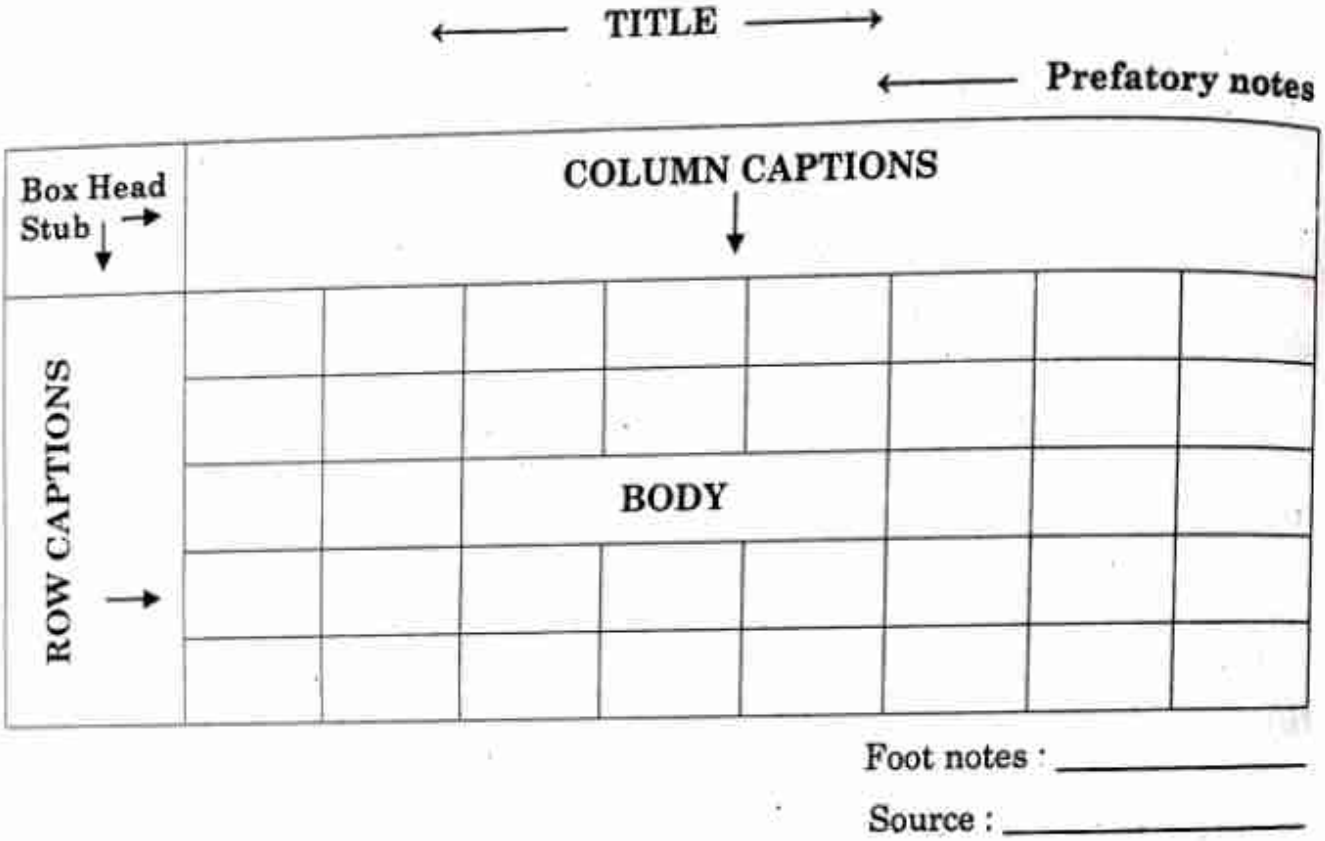
(vi) Foot Notes

It is given below the body of the table. Footnotes are used to clarify any information in the table. It provides also additional information about the table.

(vii) Source Notes

The source note is given at the end of the table. It indicates the source from where information have been taken. It includes the information about compiling agency, publication and page number etc.

GENERAL SKETCH OF A TABLE



2.3.2 General Rules of Tabulation

(i) A table should be simple and easy to understand.

(ii) A table should be complete and self-explanatory.

(iii) The row and column captions should be arranged in a systematic manner.

(iv) Do not use ditto marks. If a figure is repeated show it each time.

(v) Abbreviations should be avoided, especially in the title and headings.

2.3.3 Types of Tabulation

1. One way Tabulation

When the data are tabulated according to one characteristic, then it is called one way tabulation. *e.g.* Division wise population of Punjab according to 1998 census.

Division Wise Population of Punjab According to 1998 Census

DIVISION	Population (in thousands)
Bahawalpur	7518
D.G. Khan	6402
Faisalabad	9735
Gujranwala	11115
Lahore	13985
Multan	11428
Rawalpindi	6780
Sargodha	5622

Source : Population Census Report, 1998.

Two Way Tabulation

When the data are tabulated according to two characteristics / criteria, then it is called two way tabulation. *e.g.*, Division wise population of Punjab by Sex.

Division Wise Population of Punjab by Sex (1998 Census)

Division	Population (in thousands)		
	Male	Female	Total
Bahawalpur	3918	3600	7518
D.G. Khan	3330	3072	6402
Faisalabad	5053	4682	9735
Gujranwala	5694	5421	11115
Lahore	7321	6664	13985
Multan	5935	5493	11428
Rawalpindi	3396	3384	6780
Sargodha	2863	2759	5622

Source : Population Census Report, 1998.

Three Way Tabulation

When the data are tabulated according to three characteristics at a time, then it is called three-way tabulation.

2.4 FREQUENCY DISTRIBUTION

The arrangement of data according to size and magnitude is called frequency distribution.

Group Data

Data presented in the form of frequency distribution are called group data.

Ungroup Data

The data, which are collected for the first time and original in character, are called ungroup data.

2.4.1 Formation of Frequency Distribution

There are two types of frequency distribution.

1. Discrete Frequency Distribution
2. Continuous Frequency Distribution

2.4.2 Formation of Discrete Frequency Distribution

To prepare the discrete frequency distribution following steps are taken.

(i) Find the largest and smallest value of the given data.

(ii) Place all possible values of the variable from the smallest to largest in the first column with the title of the variable.

(iii) The second column is the tally column. In this column a vertical bar called tally bar is put against the particular value to which it relates.

(iv) Last column is the frequency column. Count all the bars and write these numbers in the frequency column. The total of the frequency column must be equal to the total number of observations.

When the discrete data are sufficiently large, then they are treated in the same way as continuous frequency distribution.

EXAMPLE 2.1

The numbers of children per family of 25 families are given below.

2, 3, 0, 5, 4, 1, 2, 4, 6, 3, 3, 5, 6, 4, 7, 1, 4, 3, 5, 3, 1, 2, 3, 5, 4

Represent the data in the form of discrete frequency distribution.

SOLUTION

Frequency Distribution of No. of Children

No. of Children "x"	Tally bar	No. of families "f"
0	|	1
1	|||	3
2	|||	3
3	𝍸|	6
4	𝍸	5
5	||||	4
6	||	2
7	|	1
Total	—	**25**

Terms Related with Continuous Frequency Distribution

Class Limits

The values of the classes are called class limits. Every class has two values. The smaller value is the lower class limit and the larger value is the upper class limit.

Mid Point or Class Mark

The average value of the lower and upper limit is called mid point or class mark.

$$\text{Mid points} = \frac{\text{Lower class limit} + \text{upper class limit}}{2}$$

Class Boundaries

The true or precise values, which describe the actual or true class limits of a class are called class boundaries. The upper class boundary of a class coincide with the lower class boundary of the next class.

Class Interval

The width of the class is called the class interval. It is also defined as the difference between the upper class boundary and the lower class boundary. It is denoted by h.

Class Frequency

The number of values in any class is the class frequency of that class. It is denoted by f.

2.4.3 Construction of Continuous Frequency Distribution

Following steps are involved in the construction of continuous frequency distribution.

1. Determine the Range

First of all find the range of the data. Range is the difference between the largest and the smallest value in the data.

2. Decide the Number of Classes

There is no hard and fast rule for this purpose. Number of classes should neither be very large nor very small. When the data are sufficiently large, the number of classes should lie between 5 and 20. H.A sturges has given a formula for determining the number of classes *i.e.*,

$$K = 1 + 3.3 \log N$$

Where

K = Number of classes

N = Total No. of observations

For example if there are 50 observations, then

$$K = 1 + 3.3 \log 50$$

$$K = 1 + 3.3 (1.6990)$$

$$= 1 + 5.607$$

$$K = 6.6067 \approx 7$$

i.e. 7 classes

But this rule is rarely used in practice.

3. Decide the size of the Class Interval

The size of the class interval depends upon the range and the number of classes. Sturges has also given a formula for this purpose.

$$\text{Class interval} = \frac{\text{Range}}{K}$$

In case of fractional results, the next higher whole number is usually taken as the size of the class interval.

4. Deciding the Starting Point

The lowest class usually starts with a number that is a multiple of class interval such that it should cover the smallest value of the data.

5. Determine the Remaining Class Limits

The remaining lower and upper class limits may be obtained by adding the size of class interval.

6. Distribute the data into Appropriate Classes

All the observations are put into respective classes. This may be done by using "Tally Bar". The frequency column is obtained by counting the tally bars of each class.

7. Total of the Frequency Column

The total of the frequency column must be equal to the total number of observations to see that all the data have been accounting for.

EXAMPLE 2.2

Marks obtained by 60 students of a class are given below.

60, 50, 46, 28, 58, 64, 36, 20, 50, 18, 42, 56, 20, 38, 40, 34, 24, 64, 64, 42, 46, 52, 50, 44, 36, 0, 24, 30, 46, 40, 64, 40, 36, 14, 36, 8, 56, 40, 30, 36, 24, 22, 36, 50, 58, 16, 40, 34, 0, 42, 42, 0, 36, 18, 18, 68, 30, 46, 38, 16.

Make a frequency distribution using appropriate class interval.

SOLUTION

First of all we find the range

Range = Largest value – smallest value

= 68 – 0 Largest value = 68

= 68 Smallest value = 0

Now we determine the No. of classes.

$K = 1 + 3.3 \log N$

$K = 1 + 3.3 \log (60)$

$= 1 + 3.3 (1.7782)$

$= 1 + 5.87 \Rightarrow K = 6.87 \approx 7$ classes

Now we decide the size of the class interval

$$h = \frac{\text{Range}}{K} = \frac{68}{7} = 9.71 \quad \textit{i.e.,}\ h = 10$$

Size of class interval = 10

Frequency distribution of the marks of students

Marks	Tally Bars	Frequency
0 – 9	𝍸	4
10 – 19	𝍸𝍸	6
20 – 29	𝍸𝍸 ||	7
30 – 39	𝍸𝍸 𝍸𝍸 ||||	14
40 – 49	𝍸𝍸 𝍸𝍸 ||||	14
50 – 59	𝍸𝍸 ||||	9
60 – 69	𝍸𝍸 |	6
Total	--	60

EXAMPLE 2.3

Form a frequency distribution from the following data taking 0.05 as the class interval and 1.19 as the lowest class limit.

1.35, 1.46, 1.50, 1.32, 1.45, 1.24, 1.49, 1.64, 1.47, 1.59, 1.41, 1.48, 1.36, 1.48, 1.51, 1.45, 1.26, 1.38, 1.76, 1.63, 1.19, 1.56, 1.65, 1.54, 1.61, 1.73, 1.60, 1.50, 1.45, 1.76, 1.67, 1.35, 1.55, 1.68, 1.46, 1.40, 1.32, 1.47, 1.64, 1.45

SOLUTION

FREQUENCY DISTRIBUTION

Class Limits	Tally Bars	Frequency (f)
1.19 – 1.23	|	1
1.24 – 1.28	||	2
1.29 – 1.33	||	2
1.34 – 1.38	||||	4
1.39 – 1.43	||	2
1.44 – 1.48	𝍸 𝍸	10
1.49 – 1.53	||||	4
1.54 – 1.58	|||	3
1.59 – 1.63	||||	4
1.64 – 1.68	𝍸	5
1.69 – 1.73	|	1
1.74 – 1.78	||	2
Total	--	**40**

Cumulative Frequency

The total frequency of all classes less than the upper class boundary of a given class is called cumulative frequency of that class.

Less than Cumulative Frequency

If the Cumulative process is from the lowest value to the highest, it is called less than cumulative frequency.

More than Cumulative Frequency

If the cumulative process is from the highest value to the lowest value, it is called more than cumulative frequency.

Relative frequency

The frequency of a class divided by the total frequency is called relative frequency of that class.

$$\text{Relative frequency} = \frac{\text{Class frequency}}{\text{Total frequency}}$$

EXAMPLE 2.4

The following table gives the frequency distribution of marks obtained by 60 students of a class. (See Example 2.2)

Marks	0 – 9	10–19	20–29	30–39	40–49	50–59	60–69
No. of Students	4	6	7	14	14	9	6

Construct the following.

(i) Class boundaries and mid points.

(ii) A less than cumulative frequency.

(iii) A more than cumulative frequency.

(iv) Relative frequency.

SOLUTION

(i) **Class boundaries and Mid Points**

Class Limits	Frequency	Class Boundaries	Mid Point
0 – 9	4	0 – 9.5	4.5
10 – 19	6	9.5 – 19.5	14.5
20 – 29	7	19.5 – 29.5	24.5
30 – 39	14	29.5 – 39.5	34.5
40 – 49	14	39.5 – 49.5	44.5
50 – 59	9	49.5 – 59.5	54.5
60 – 69	6	59.5 – 69.5	64.5

(ii) Less than Cumulative Frequency

End values (Upper C.B.)	f	Cumulative Frequency
Less than 0	0	0
Less than 9.5	4	4
Less than 19.5	6	4 + 6 = 10
Less than 29.5	7	10 + 7 = 17
Less than 39.5	14	17 + 14 = 31
Less than 49.5	14	31 + 14 = 45
Less than 59.5	9	45 + 9 = 54
Less than 69.5	6	54 + 6 = 60

(iii) More than Cumulative Frequency

End values Lower C.B.)	f	Cumulative Frequency
More than 0	4	56 + 4 = 60
More than 9.5	6	50 + 6 = 56
More than 19.5	7	43 + 7 = 50
More than 29.5	14	29 + 14 = 43
More than 39.5	14	15 + 14 = 29
More than 49.5	9	6 + 9 = 15
More than 59.5	6	6
More than 69.9	0	0

(iv) **Relative Frequency**

Marks	f	Relative Frequency
0 – 9	4	$\frac{4}{60} = 0.07$
10 – 19	6	$\frac{6}{60} = 0.10$
20 – 29	7	$\frac{7}{60} = 0.12$
30 – 39	14	$\frac{14}{60} = 0.23$
40 – 49	14	$\frac{14}{60} = 0.23$
50 – 59	9	$\frac{9}{60} = 0.15$
60 – 69	6	$\frac{6}{60} = 0.10$
Total	60	1.00

2.5 GRAPHIC AND DIAGRAMMATIC REPRESENTATION OF DATA

Visual representation of data in the form of points, lines, symbols and areas are called Grapnic representation. Such visual representation of data can be divided into two main groups namely diagrams and graphs.

2.5.1 Diagrams

There are different types of diagrams. Commonly used diagrams are

(a) One dimensional diagram

(b) Aerial or two dimensional diagrams

(c) Pie Diagrams

2.6 ONE DIMENSIONAL DIAGRAMS

These are the diagrams in which only the length is taken into account but are of equal width. One dimensional diagrams are further divided into

(i) Simple Bar Diagram

(ii) Multiple Bar Diagram

(iii) Sub divided or Component Bar Diagrams

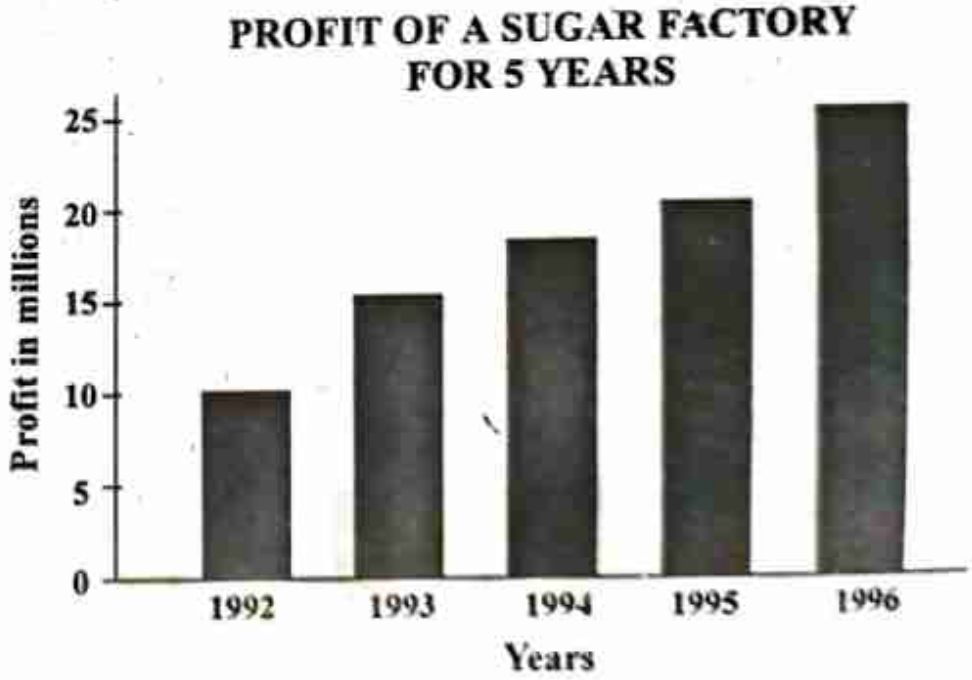
2.6.1 Simple Bar Diagrams

A simple bar diagram is used when the data consist of single factor, such as production, profit, yield etc. It consists of horizontal or vertical bars of equal width The length of the bar is in proportion to the actual data. The space between the bars should not exceed than the width of the bar. The bars should neither very tall nor very small. The data should be arranged in ascending or descending order if the time is not given.

EXAMPLE 2.5

Draw a simple bar diagram to represent the profits of a sugar factory for 5 years.

Year	1992	1993	1994	1995	1996
Profit (Million Rs.)	10	15	18	20	25



2.6.2 Multiple Bar Diagram

Multiple bar diagram is an extension of simple bar diagram, which is applicable to only one factor. But in multiple bar diagrams two or more factors are

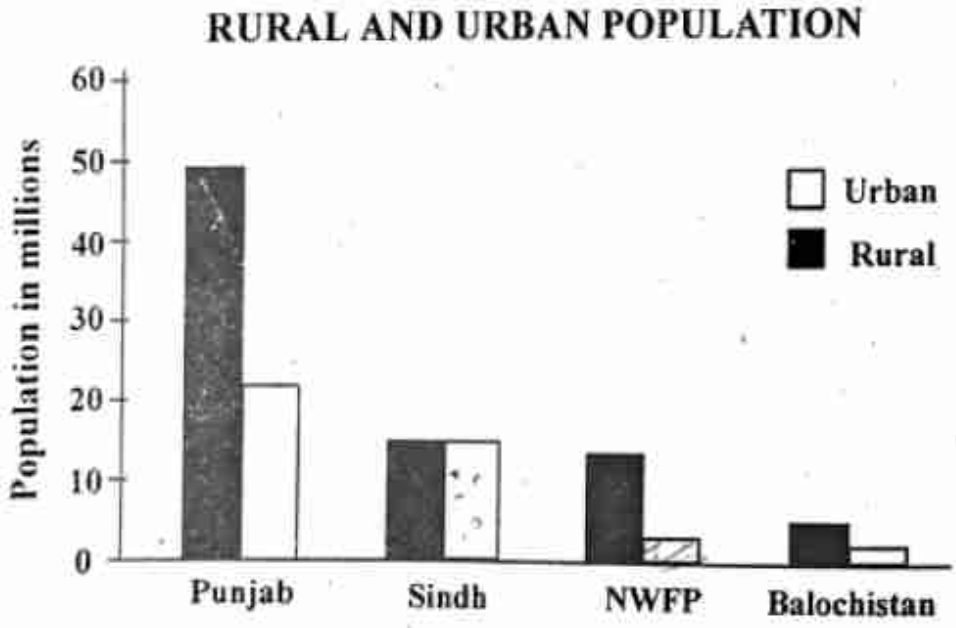
taken into account. In this diagram, grouped bars are used to represent related sets of data. Each bar is shaded or coloured differently for identification. It is a good device for the comparison of two or more than two kind of information. For example, exports and imports, birth and death rates of a country can be compared.

EXAMPLE 2.6

Draw a multiple bar diagram to represent the Rural and Urban population of four provinces of Pakistan in 1998.

Provinces	Population in Millions	
	Rural	Urban
Punjab	49.9	22.7
Sindh	15.3	14.7
N.W.F.P	14.6	3.0
Balochistan	5.0	1.5

Source : Population Census Report, 1998.



2.6.3 Component Bar Diagram

Component bar diagrams are those in which each bar represent the total of the components. Then it is divided into two or more sections in the ratio of their

components. The components parts are coloured or shaded differently for identification. They are also known as sub-divided bars.

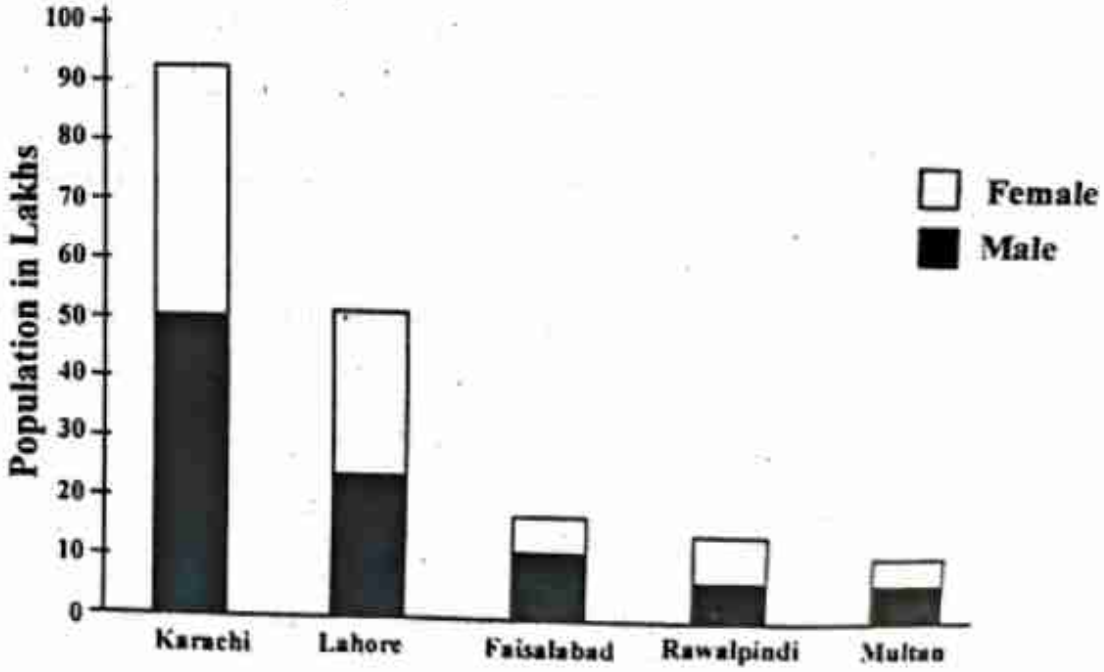
EXAMPLE 2.7

Draw a component bar diagram to represent the male and female population of 5 big cities in Pakistan in 1998.

Cities	Population in Lakhs		
	Male	Female	Total
Karachi	50	43	93
Lahore	26.6	24	50.6
Faisalabad	10.4	9.4	19.8
Rawalpindi	7.5	6.6	14.1
Multan	6.3	5.5	11.8

Source : Pakistan Census Report, Govt. of Pakistan, July 1998.

MALE AND FEMALE POPULATION OF FIVE BIG CITIES IN PAKISTAN IN 1998



2.7 AERIAL OR TWO DIMENSIONAL DIAGRAMS

Two dimensional diagrams are based on the area. The area is the product of length and breadth. Hence in two dimensional diagrams, both length and breadth are taken into account. They are in the form of rectangles, subdivided rectangles and squares.

2.7.1. Rectangular Diagrams

When the data have moderately large variations, rectangular diagrams can be used. The area of the rectangle is equal to the product of its length and breadth. There are two methods of drawing rectangles.

(i) The lengths of the rectangle are kept constant and breadths are taken proportional to the size of data.

(ii) Secondly breadths are kept constant and length are taken proportional to the size of data.

EXAMPLE 2.8

Draw rectangles to represent the are of four provinces of Pakistan.

Provinces	Area (Thousand square km)
Balochistan	347
Punjab	205
Sindh	141
K.P.K	75

SOLUTION

In order to keep the breadths constant. We split the total into two portions as

Area of Balochistan $= 347 = 10 \times 34.7$

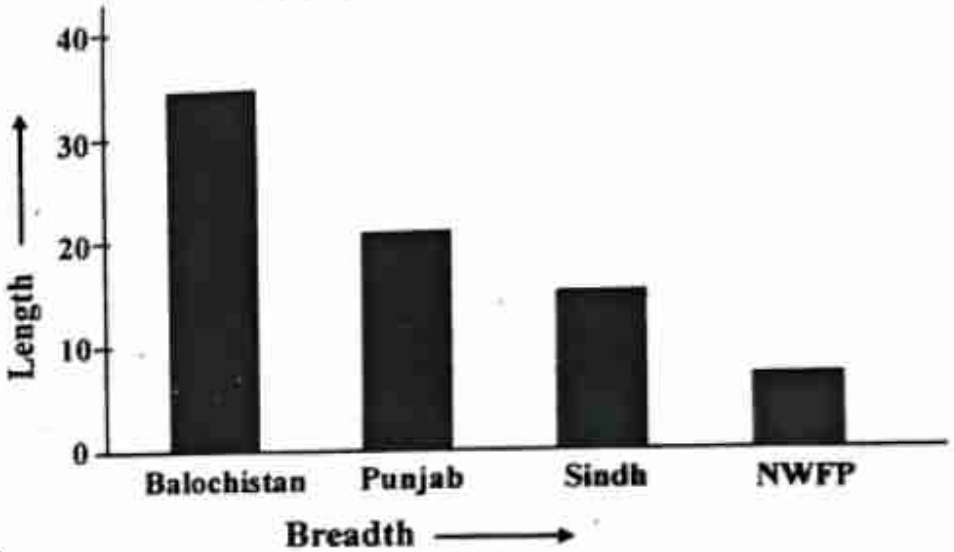
Area of Punjab $= 205 = 10 \times 20.5$

Area of Sindh $= 14.1 = 10 \times 14.1$

Area of K.P.K $= 75 = 10 \times 7.5$

In this question, the breadth of each rectangle is equal to 10 unit, while the length of rectangles are 34.7, 20.5, 14.1 and 7.5 units for Balochistan, Punjab, Sindh and KPK respectively.

RECTANGULAR DIAGRAM



2.7.2. Sub-divided Rectangles

Sub divided rectangular diagrams is used to represent the data where the quantities along with their components are to be compared. These diagrams are generally used to compare the budgets of different families.

In this diagram, rectangles are drawn with length equal to 100 units and breadth proportional to the size of the values. The component parts are expressed as percentage of their corresponding totals. Each rectangle is divided into different sections. The component parts are coloured or shaded differently to increase the effectiveness of the diagram.

EXAMPLE 2.9

Draw a sub-divided rectangular diagram to compare the budgets of two families A and B.

Items of Expenditure	Family A	Family B
Food	480	1200
Clothing	240	850
Fuel & Lighting	80	250
Housing	180	600
Service	100	400
Misc.	120	300

SOLUTION

The necessary calculations are given below.

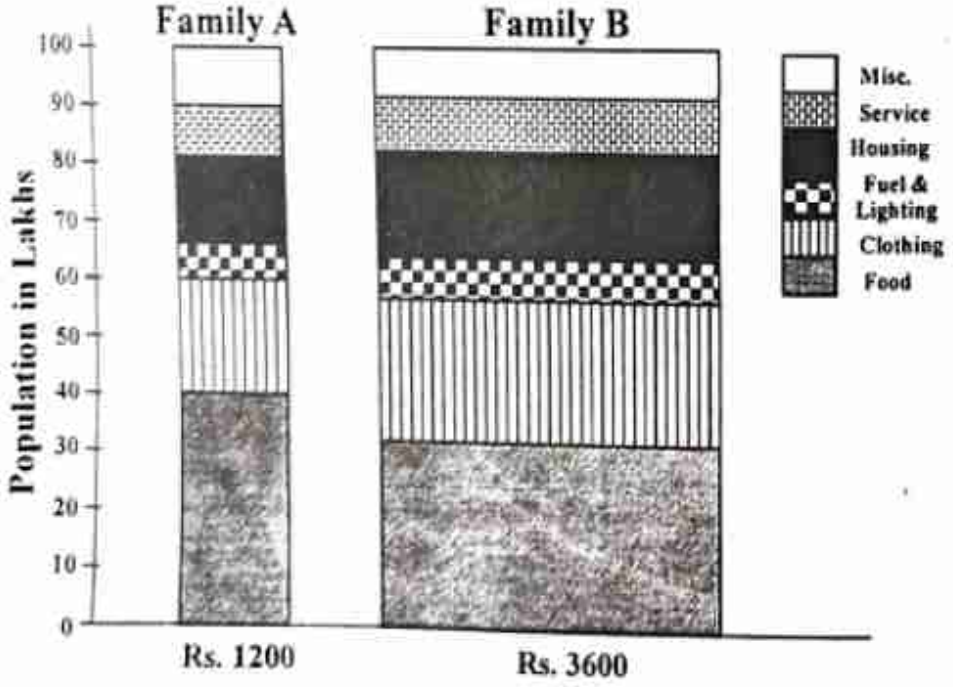
Items of Expenditure	Family A			Family B		
	Expend	% Expend	Cum %	Expend	% Expend	Cum %
Food	480	40.00	40.00	1200	33.33	33.33
Clothing	240	20.00	60.00	850	23.61	56.94
Fuel & Light	80	6.67	66.67	250	6.94	63.88
Housing	180	15.00	81.67	600	16.67	80.55
Service	100	8.33	90.00	400	11.11	91.66
Misc.	120	10.00	100.00	300	8.34	100.00
Total	1200	100.00	—	3600	100.00	—

Length of the rectangles = 100 units

The breadth of the rectangles are in proportion = 1200 : 3600

i.e 1 : 3

SUB-DIVIDED RECTANGULAR DIAGRAM



2.7.3 Pie Diagrams

Like the component bar diagram, pie diagrams show the relationship between the whole and its components. However, there is an important difference. In component bar diagrams, the length of bars are compared, where as in pie diagrams, areas of the sectors are compared.

To construct a pie diagram, draw a circle of any suitable radius. A circle consists of 360°. To show the component parts by sectors, we calculate the angles for each sector by using the relation.

$$\text{Angle} = \frac{\text{Component Part}}{\text{Whole quantity}} \times 360$$

Then divide the circle into different sectors by constructing angles at the center by means of a protractor. The different sectors are shaded differently for identification.

EXAMPLE 2.10

Draw a pie diagram for the following data.

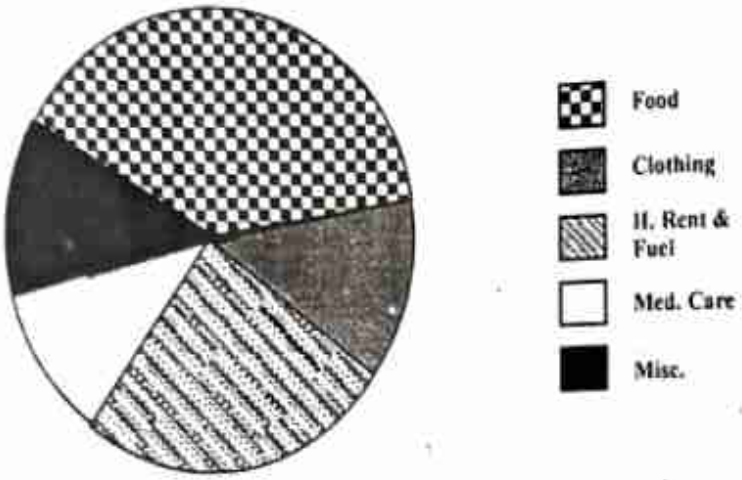
Items	Food	Clothing	H. Rent & Fuel	Med. Care	Misc.
Exp.	90	30	50	40	30

SOLUTION

The necessary calculations are given below.

Items	Expenditure (in Rs.)	Angle of Sectors (in Degrees)
Food	90	$\frac{90}{240} \times 360 = 135$
Clothing	30	$\frac{30}{240} \times 360 = 45$
H. Rent & Fuel	50	$\frac{50}{240} \times 360 = 75$
Medical Care	40	$\frac{40}{240} \times 360 = 60$
Misc.	30	$\frac{30}{240} \times 360 = 45$
Total	240	360

PIE DIAGRAM



2.8 GRAPHIC REPRESENTATION

Graphic representation means the visual representation of the numerical data. The movement of data can be presented very effectively by means of a graph. It is the simplest and most widely used method of presenting the data. A graph shows the relationship between two variables, for example, the amount of sales and the period of time when the sales were made. Graphs provide an overall picture of a statistical series. Graphs can also be used to make predictions and forecast. But graphs are less accurate and convey limited information. That is the only disadvantage of a graph.

2.8.1 Construction of Graphs

The first step in the construction of a graph is the drawing of two lines at right angles. The hoizontal line is called x-axis or Abscissa and the vertical line is known as y-axis or ordinate. These two lines together are known as co-ordinate axis. the point of inter section of the two axis is called the origin which is denoted by O. Some suitable scales are selected along x-axis and y-axis independent variable is taken along x-axis and dependent variable along y-axis. Points are plotted along both axis and joined to get the required graph.

2.9 TYPES OF GRAPHS

Graphs can be divided into two main groups, namely;

(a) Graphs of Time Series (Historigram)

(b) Graphs of Frequency distribution(Histogram)

2.9.1 Graph of Time Series (Historigram)

A curve showing the changes in the value of one or more than one items over a period of time is called the graph of a time series. A graph of time series is called

Historigram. The time is taken along x-axis and the other variable is along y-axis. Points are plotted and joined by means of a straight line to get the historigram.

EXAMPLE 2.11

The price of wheat per 40 kg during the years 1993 to 2000 are given below. Draw a Historigram.

Years	1993	1994	1995	1996	1997	1998	1999	2000
Price	130	160	160	173	240	240	240	300

2.9.2 Graphs of Frequency Distribution

The important graphs of frequency distribution are

(i) Histogram

(ii) Frequency Polygon

(iii) Frequency Curves

(iv) Cumulative frequency polygon (OGIVE)

2.9.3 Histogram

A frequency distribution can be shown in the form of a diagram. To construct histogram class boundaries are taken along x-axis and the frequency along y-axis. When the class intervals are equal, the rectangles are completed with class interval as the width and heights proportional to the frequencies. The rectangles are adjacent to each other and there is no gap between the rectangles. It is used to locate mode graphically.

If the class intervals are unequal, then the heights of the rectangles are adjusted. Each class frequency is divided by its interval size, then rectangles are completed with class interval as a width and heights proportional to the ratio of the frequencies.

EXAMPLE 2.12

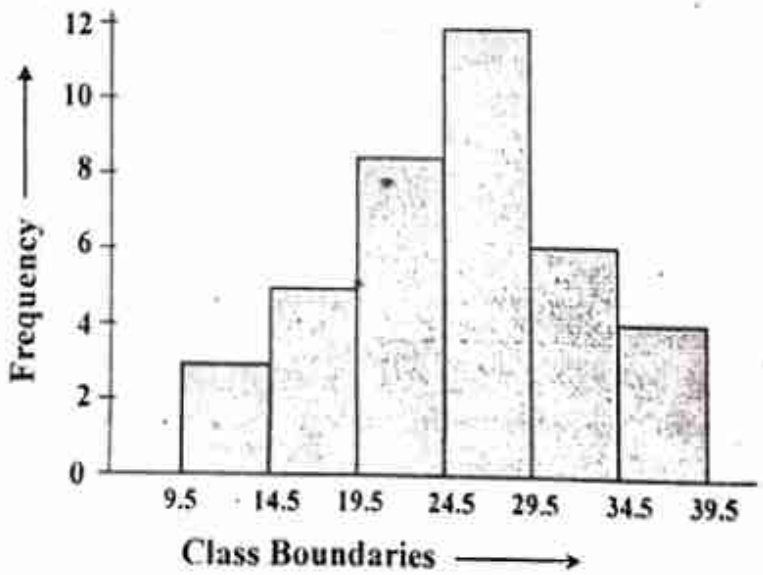
Construct Histogram for the following frequency distribution.

Groups	10 – 14	15 – 19	20 – 24	25 – 29	30 – 34	35 – 39
Freq.	3	5	9	12	6	4

SOLUTION

Groups	f	Class Boundaries
10 – 14	3	9.5 – 14.5
15 – 19	5	14.5 – 19.5
20 – 24	9	19.5 – 24.5
25 – 29	12	24.5 – 29.5
30 – 34	6	29.5 – 34.5
35 – 39	4	34.5 – 39.5

HISTOGRAM



EXAMPLE 2.13

Construct a Histogram for the following frequency distribution.

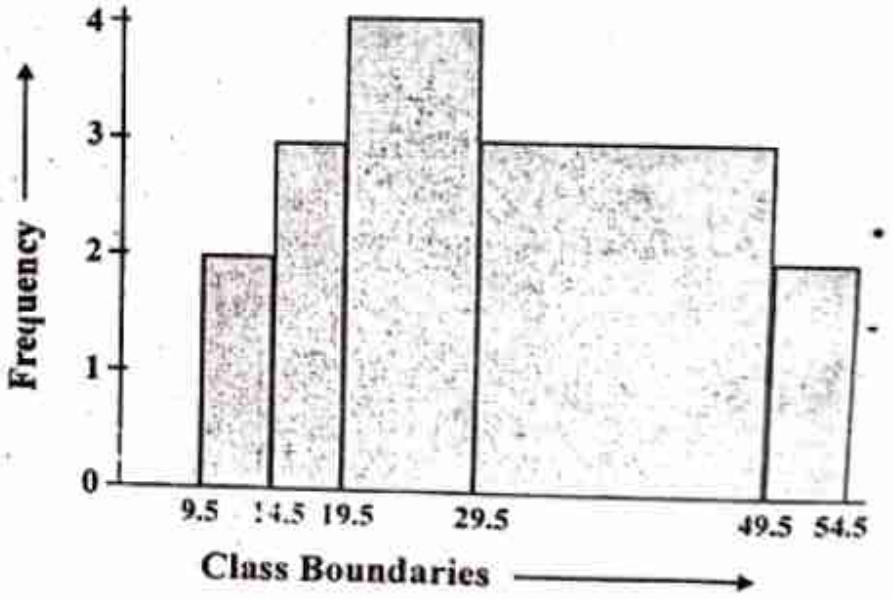
Classes	10–14	15–19	20–29	30–49	50–54
Frequency	10	15	40	60	10

SOLUTION

Here the class interval are unequal. The heights of the rectangles may be adjusted. Each class frequency is divided by its class interval size. The necessary calculations are given below.

Classes	Class interval	f	Proportional height	Class Boundaries
10 – 14	5	10	$\frac{10}{5} = 2$	9.5 – 14.5
15 – 19	5	15	$\frac{15}{5} = 3$	14.5 – 19.5
20 – 29	10	40	$\frac{40}{10} = 4$	19.5 – 29.5
30 – 49	20	60	$\frac{60}{20} = 3$	29.5 – 49.5
50 – 54	5	10	$\frac{10}{5} = 2$	49.5 – 54.5

HISTOGRAM FOR UNEQUAL CLASS INTERVAL



2.9.4 Frequency Polygon

Frequency polygon is obtained by joining the mid points of the tops of the rectangles of a histogram by means of a straight line. As we know that polygon is a many sided closed figure, therefore, we add extra classes at both ends of the frequency distribution with zero frequencies. By doing so, the polygon forms a closed figure.

EXAMPLE 2.14

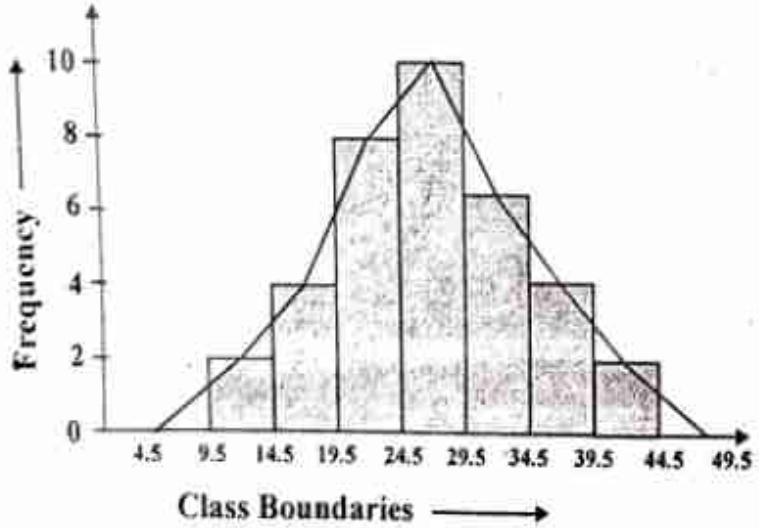
Draw a frequency polygon for the following frequency distribution.

Groups	10–14	15–19	20–24	25–29	30–34	35–39	40–44
Freq.	2	4	8	10	7	4	2

SOLUTION

Groups	f	Class boundaries
10 – 14	2	9.5 – 14.5
15 – 19	4	14.5 – 19.5
20 – 24	8	19.5 – 24.5
25 – 29	10	24.5 – 29.5
30 – 34	7	29.5 – 34.5
35 – 39	4	34.5 – 39.5
40 – 44	2	39.5 – 44.5

FREQUENCY POLYGON



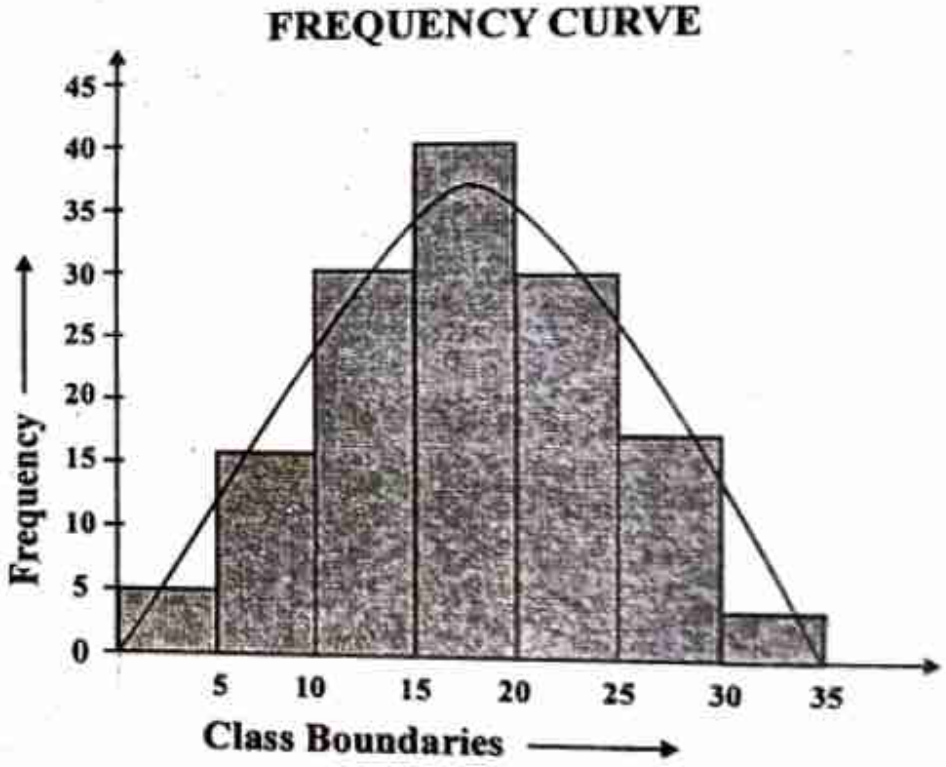
2.9.5 Frequency Curves

If the class intervals are very small and the number of classes is very large in a frequency distribution, then the histogram and frequency polygon approach a smooth curve *i.e.* there is no ups and downs in the curve. That smooth curve is called frequency curve. In the frequency curve the mid points of tops of rectangles of histogram are not joined together by straight lines. The free hand drawing method of drawing curve is used.

EXAMPLE 2.15

Draw a frequency curve of the following data.

Marks	0–5	5–10	10–15	15–20	20–25	25–30	30–35
No. of Students.	5	16	30	40	30	18	4



2.9.6 Cumulative Frequency Polygon (Ogive)

In cumulative frequency polygon, the cumulative frequencies are plotted against the upper class boundaries. The upper class boundaries are taken along x-axis and the cumulative frequency along y-axis. The plotted points are joined by a straight line. This curve is known as cumulative frequency polygon. It is also called Ogive. It is used to locate median, Quartiles, Deciles and Percentiles.

EXAMPLE 2.16

Draw cumulative frequency polygon or Ogive from the following data.

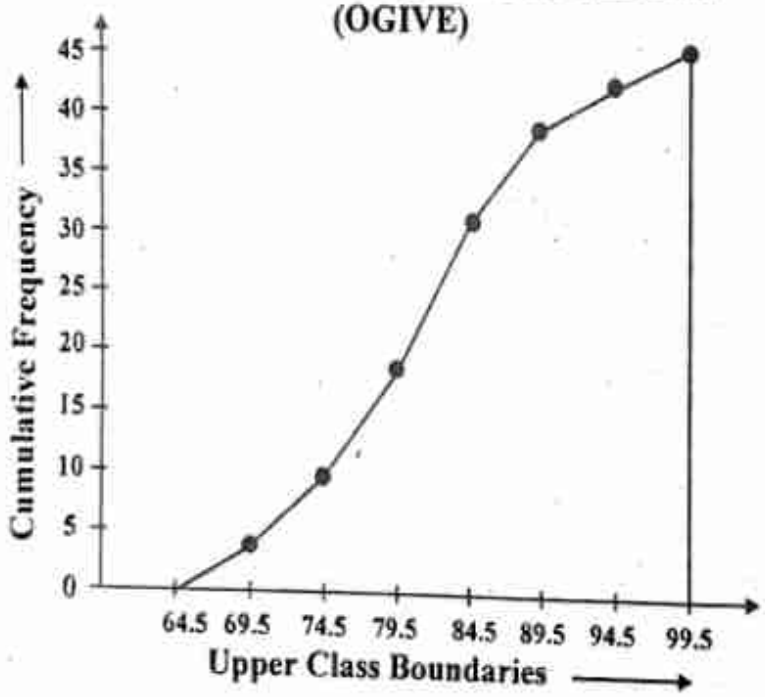
Daily wages	65 - 69	70 - 74	75 - 79	80 - 84	85 - 89	90 - 94	95 - 99
No. of Persons	4	6	9	12	8	4	2

SOLUTION

The necessary calculations are given below.

Classes	Class Boundaries	f	C.f.	U.C.B.
65 – 69	64.5 – 69.5	4	4	69.5
70 – 74	69.5 – 74.5	6	10	74.5
75 – 79	74.5 – 79.5	9	19	79.5
80 – 84	79.5 – 84.5	12	31	84.5
85 – 89	84.5 – 89.5	8	39	89.5
90 – 94	89.5 – 94.5	4	43	94.5
95 – 99	94.5 – 99.5	2	45	99.5

CUMULATIVE FREQUENCY POLYGON (OGIVE)



Difference between Histogram and Historigram

Histogram is the graph of frequency distribution, while Historigram is the graph displaying the variation in the time series dealing with prices, production, export and import etc.

EXERCISES

2.1 What is meant by classification? Write down the basis of classification and explain it.

2.2 Define the term tabulation. Discuss the main parts of a table.

2.3 (a) Differentiate between classification and tabulation.

(b) Distinguish between simple and compound tables.

2.4 **Define the following terms.**

(i) Frequency distribution

(ii) Class-interval

(iii) Class limits

(iv) Class marks / mid points

(v) Class boundaries

2.5 (a) What is meant by frequency distribution. Describe briefly the main steps involved in the preparation of frequency distribution.

(b) The following data shows the number of road side accidents per day.

2,5,1,2,0,6,3,4,5.1,2,5,3,0,6,4,0,6,3,2,4,3,6,7,2

Make a discrete frequency distribution.

The number of Persons in 20 families are given below:

2, 1, 3, 2, 4, 2, 4, 5, 4, 5, 3, 4, 5, 5, 4, 5, 4, 5, 4, 5.

Construct an ungrouped frequency distribution.

2.6 The number of children born to 30 women are given below.

5,3,3,2,0,6,1,4,1,6,3,4,8,7,2,4,5,2,4,6,4,3,3,0,5,2,1,4,5,3

Represent the data in the form of frequency distribution taking one as a class interval.

2.7. The weights in pounds of 30 college students are given below.

130, 133, 124, 121, 115, 139, 137, 144, 142, 133, 133, 128, 129, 132, 131, 128, 126, 132, 134, 135, 138, 130, 141, 136, 135, 141, 123, 126, 118, 134

Prepare a frequency distribution taking a class interval of size 5.

2.8 Tabulate the following marks into frequency distribution taking 10 as the class interval and 45 as the lower limit.

109, 74, 49, 103, 95, 90, 118, 52, 88, 101, 96, 72, 56, 64, 110, 97, 59, 62, 96, 82, 65, 85, 105, 116, 91, 83, 99, 52, 76, 84, 89, 77, 104

2.9 The following data give the index number of commodities in a certain year. Form a frequency distribution using suitable class interval.

70, 99, 114, 110, 88, 108, 87, 101, 71, 117, 110, 153, 145, 127, 98, 156, 127, 132, 113, 115, 100, 117, 73, 134, 108, 102, 123, 106, 119, 104, 129, 117, 106, 119, 116

2.10 Marks of students in English out of 100 are given below.

62, 59, 58, 42, 35, 60, 5, 12, 29, 52, 42, 38, 39, 45, 39, 28, 35, 37, 41, 51, 37, 39, 45, 16, 9, 17, 42, 59, 0, 23, 45, 33, 12, 40, 60

Construct

(i) A frequency distribution by using a class interval of 5.

(ii) A less than cumulative frequency.

(iii) A more than cumulative frequency.

(iv) Relative frequency.

2.11 (a) In an experiment measuring the percentage of shrinkage on dying 40 plastic clay test specimens gave the following results.

19.3, 16.9, 17.9, 17.3, 15.8, 18.5, 17.1, 19.5, 20.4, 18.7, 22.3, 17.5, 18.4, 13.9, 18.8, 16.8, 14.9, 19.5, 19.4, 16.3, 17.8, 23.4, 17.4, 19.0, 21.8, 18.8, 18.5, 18.2, 16.1, 18.3, 17.5, 17.4, 18.6, 16.9, 16.5, 18.2, 20.5, 20.5, 17.5, 19.1

Group these data into frequency distribution taking 1.00 as class interval. e.g. 13.5 – 14.4, 14.5 – 15.4 etc.

(b) Calculate class boundaries, mid point cumulative frequency and relative frequency of part (a) in question 2.11.

2.12 (a) What do you understand by diagrammatic representation of statistical data?

(b) Explain the following diagrams.

(i) Simple Bar Diagrams

(ii) Multiple Bar Diagrams

(iii) Sub-divided Rectangles

(iv) Pie Diagrams

2.13 The population of 5 big cities in Pakistan is given below. Draw a simple bar diagram.

Cities	Karachi	Lahore	Faisalabad	Rawalpindi	Multan
Pop. (in lakhs)	92.7	50.6	19.8	14.1	11.8

Source : Population Census Report, 1998.

2.14 **(a)** What is meant by tabulation.

(b) For the following data construct simple bar chart / diagram.

Years	1962	1963	1964	1965	1966	1967	1968	1969
Production.	1050	1200	1250	1370	1450	1500	1585	1620

2.15 **(a)** Male and female population of four provinces of Pakistan in 1998 are given below.

Provinces	Population (in millions)		
	Males	Female	Total
Punjab	37.5	35.1	72.6
Sindh	15.8	14.2	30.0
N.W.F.P	9.0	8.6	17.6
Balochistan	3.5	3.0	6.5

Construct

(i) Component Bar Diagram

(ii) Multiple Bar Diagram

(b) Draw a component bar chart for the following data:

(Population in Lakhs)

Division	Both Sexes	Male	Female
Peshawar	64	33	31
Rawalpindi	40	21	19
Sargodha	60	32	28
Lahore	65	35	30

2.16 (a) Draw sub-divided rectangles to compare the budgets of two families A and B.

Items of Expenditure	Family - A	Family - B
Food	240	600
Clothing	120	425
Fuel	40	125
Housing	90	300
Services	50	200
Misc.	60	150

(b) The following table gives the monthly budgets of two families. Represents the data by sub-divided rectangles.

Item of Expenditure	Family A	Family B
Food	6000	8000
Clothing	1000	1000
House Rent	4000	5000
Fuel & Lighting	1000	1000
Miscellaneous	3000	5000
Total	Rs. 15000	Rs. 20000

2.17 (a) Draw a Pie Diagram for the following data.

Items	Expenditure in Rs.
Food	95
Clothing	32
House Rent	50
Medical care	13
Misc.	50

(b) Represent the following data by a pie diagram.

Items	Expenditure in Rs.
Food	75
Clothing	50
House rent	30
Education	25
Miscellaneous	20

2.18 (a) What do you meant by graphic representation of data?

(b) Write down the various types of graphs. Explain it briefly.

2.19. (a) What is histogram. How does it differ from historigram?

(b) The heights of the college students are given below.

Heights (in inches)	57–59	60–62	63–65	66–68	69–71	72–74
No. of Students	8	15	27	18	9	3

Draw a histogram and frequency polygon.

2.20 (a) Explain the method of constructing histograms when the class intervals are unequal.

(b) Draw a histogram from the following data.

Classes	12–14	15–19	20–29	30–34	35–38	39–40
Freq.	12	25	70	25	16	6

2.21 (a) What is meant by classification.

(b) Prepare a frequency distribution from the following data taking class interval as

2.2 – 2.7, 2.8 – 3.3, ...

4.1, 3.5, 3.2, 4.2, 3.6, 3.5, 4.2, 4.8, 4.1, 4.3, 4.4, 4.3, 3.8, 4.2, 4.7, 2.8, 3.7, 4.9, 5.6, 3.3, 3.7, 2.6, 2.7, 4.7, 4.1, 4.2, 4.6, 4.1, 4.9, 3.7, 4.5, 3.7

(c) Draw a histogram for the distribution you obtain in part (b) above.

2.22 (a) Differentiate between histogram and historigram.

(b) Population of Pakistan (in millions) from 1965 - 90 is given below.

Years	1965	1970	1975	1980	1985	1990
Population (in Millions)	52.6	60.6	71.0	82.6	96.2	112.4

Source : Pakistan population welfare program, Report 1995, P-56.

Draw a historigram.

(c) For the following data draw a historigram.

Year	1962	1963	1964	1965	1966	1967	1968
Production	1050	1200	1250	1370	1450	1500	1600

2.23 (a) Define frequency polygon and cumulative frequency polygon or Ogive.

(b) Daily wages of factory workers are given below.

Wages	75–79	80–84	85–89	90–94	95–99	100–104
No. of Workers	2	4	8	11	13	7

Draw a cumulative frequency polygon or Ogive.

(c) Construct an ogive from the following table.

Weight	118-126	127-135	136-144	145-153	154-162	163-171	172-180
Frequency	3	5	9	12	5	4	2

2.24 Make a frequency distribution taking classes are;1.20 – 1.49, 1.50 – 1.79 ,..... 3.20, 3.17, 2.87, 1,45, 1.49, 2.37, 2.86, 2.50, 1.67, 2.66, 3.18, 3.06, 2.56, 1.86, 1.99, 2.06, 2.45, 2.22, 3.10, 1.72, 2.04, 2.15, 2.45, 2.68, 2.75, 2.89, 1.44, 1.60, 1.54, 1.48.

2.25 Make a frequency distribution of the following marks obtained by students of a class in the subject of mathematics:

48, 88, 70, 72, 73, 49, 5, 17, 38, 40, 56, 49, 78, 98, 89, 14, 68, 87, 98, 99, 79, 49, 40, 52, 14, 34, 78, 29, 17, 83, 70, 68, 82, 55, 48, 31, 72, 5, 12, 9, 11, 48, 59, 71, 80, 72, 62, 66, 49, 48.

Taking class limits 1 – 20, 21 – 40 and so on.

2.26. SELECT THE CORRECT ANSWER

(i) The process of arranging the data into homogenous classes according to their similarities is called

(a) average (b) classification
(c) tabulation (d) data array

(ii) The basis of classification are

(a) two (b) three
(c) four (d) five

(iii) A table has at least

(a) one part (b) two parts
(c) three parts (d) four parts

(iv) In table, column caption is also called

(a) box head (b) stub
(c) body (d) title

(v) Total angle of the Pie chart is

(a) 180° (b) 270°
(c) 360° (d) 380°

(vi) In pie chart the arrangement of angles of different sectors are generally

(a) clockwise (b) anti clockwise
(c) random (d) none of these

(vii) The grouped data is also called

(a) raw data (b) primary data
(c) secondary data (d) simple values

(viii) The data related to export and import of a country should be presented by

(a) simple bar chart (b) multiple bar chart
(c) pie chart (d) aerial chart

(ix) When the classes are 1 – 3, 4 – 6, 7 – 9, ... the class interval is

(a) 2 (b) 3
(c) 4 (d) 5

(x) The total of relative frequency is always equal to

(a) 1 (b) 0.5
(c) –1 (d) 100

(xi) The number of values in any class is called

(a) mid point (b) class interval
(c) frequency (d) cumulative frequency

(xii) The difference between maximum and minimum value is called

(a) range (b) class interval

(c) mid point (d) frequency

(xiii) At G.P.O. sorting of letters is a type of

(a) one way tabulation (b) frequency distribution

(c) tabulation (d) classification

(xiv) Graph of the frequency distribution is called

(a) Histogram (b) Historigram

(c) Ogive (d) Frequency polygon

(xv) Graph of the time series is called

(a) Histogram (b) Historigram

(c) Ogive (d) Frequency polygon

(xvi) The number of observations falling in a particular class is called:

(a) Class mark (b) Class interval

(c) Mid point (d) None of these

(xvii) A pie chart is represented by:

(a) Rectangle (b) Square

(c) Circle (d) Triangle

(xviii) A sector diagram is also called:

(a) Bar diagram (b) Histogram

(c) Pie diagram (d) Component bar diagram

Answers 2.26

(i)	(b)	**(ii)**	(c)	**(iii)**	(d)	**(iv)**	(a)	**(v)**	(c)
(vi)	(b)	**(vii)**	(c)	**(viii)**	(b)	**(ix)**	(b)	**(x)**	(a)
(xi)	(c)	**(xii)**	(a)	**(xiii)**	(d)	**(xiv)**	(a)	**(xv)**	(b)
(xvi)	(d)	**(xvii)**	(c)	**(xviii)**	(c)				

SHORT QUESTION ANSWERS

1. **What is classification?**

Ans. The process of arranging the data into homogenous classes according to their resemblances and similarities is called classification *e.g.*, the process of sorting letters.

2. **What are the basis of classification?**

Ans. The basis of classification are

(a) Geographical or spatial

(b) Chronological or temporal

(c) Qualitative

(d) Quantitative

3. **Define Geographical or spatial classification.**

Ans. In geographical classification, the data are classified by geographical region or locations. For example, population of a country may be classified by provinces, divisions, districts and tehsils etc.

4. **Define chronological or temporal classification.**

Ans. When the data are classified according to the time of its occurrence such as years, months weeks, days, hours etc.

5. **Define qualitative classification.**

Ans. In qualitative classification, the data are classified on the basis of some quality or attribute, such as religion, sex, beauty, colour etc.

6. **What do you mean by quantitative classification?**

Ans. In this type data are classified according to some quantitative measurements *e.g.* height, income, age, sale etc.

7. **Define range.**

Ans. The difference between maximum and minimum value is called range.

i.e. $R = X_m - X_o$

8. **Define tabulation.**

Ans. The systematic arrangement of data into rows and columns is called tabulation.

9. What are the main parts of a table?

Ans. The main parts of the table is

(a) Title

(b) Prefatory notes

(c) Column captions and Box head

(d) Row captions and stub

(e) Body

(f) Foot notes

(g) Source notes

10. Define one way tabulation.

Ans. When the data are tabulated according to one characteristics is called one way tabulation. *e.g.* division wise population of Punjab.

11. Define two way tabulation.

Ans. When the data are tabulated according to two characteristics, then it is called two way tabulation. *e.g.* Division wise population of the Punjab by sex. (Male and Female)

12. Define frequency distribution.

Ans. The arrangement of data according to size and magnitude is called frequency distribution.

13. Write down the steps for construction of continuous frequency distribution.

Ans. Following steps are involved.

(a) Determine the range

(b) Decide the number of classes

(c) Class interval size

(d) Decide the starting point

(e) Determine the remaining class limits.

(f) Distribute the data into appropriate classes.

14. Define midpoint or class mark.

Ans. The average value of the lower and upper class limit is called midpoint or class mark.

15. What is the difference between class limits and class boundaries?

Ans. Class Limit

The values of the classes are called class limits. Every class has two values. The smaller is called lower class limit and the larger is called upper class limit.

Class Boundaries

The true or precise values, which describe the actual or true class limits of a class, are called class boundaries.

16. Define class frequency.

Ans. The number of values in any class is the class frequency of that class.

17. What do you mean by class interval?

Ans. The width of the class is called class interval. It is denoted by h.

18. Define cumulative frequency.

Ans. The total frequency of all classes less than the upper class boundary of a given class is called cumulative frequency of that class.

19. Define relative frequency.

Ans. The frequency of a class divided by the total frequency is called relative frequency of that class.

$$\text{Relative frequency} = \frac{\text{Class frequency}}{\text{Total frequency}}$$

20. What do you know about graphic representation?

Ans. Visual representation of data in the form of points, lines, symbols or areas are called Graphic representation.

21. What do you mean by pie chart?

Ans. Pie chart/diagram show the relationship between the whole and its components. In pie chart areas of the sectors are compared.

To construct a pie chart, draw a circle of suitable radius. To show the components parts by sectors, we calculate the angles for each sector by using the relation.

$$\text{Angle} = \frac{\text{Component Part}}{\text{Whole quantity}} \times 360^\circ$$

22. What do you know about cumulative frequency polygon or ogive.

Ans. It is used to locate median. In cumulative frequency polygon, the cumulative frequencies are plotted against the upper class boundaries. The upper class boundaries are taken along x-axis and the cumulative frequency along y-axis. The plotted points are joined by a straight line. This curve is known as cumulative frequency curve.

23. What is the difference between Histogram and Historigram?

Ans. Histogram is the graph of frequency distribution, while historigram is the graph of time series.