Statistics:

Statistics is the collection, presentation analysis and utilize of numeric data to make inferences and reach decision in the face of uncertainty in economics business and other social and physical science.

Branches of statistics:

Two branches of statistics

1: Theoretical/mathematical statistics

2: Applied statistics

Theoretical statistics:

It is branch of statistics that formulate statistical methods and general rules to be applied for investigation to a specific problem.

This branch of statistics is divided into further three branches.

Design experiment:

The design of experiment is the sequence of steps taken to collect appropriate data for objective analysis to draw valid inference with respect to the problem under investigation.

Descriptive statistics:

A branch of statistics in which we collect, arrange and analyze the numerical data and no conclusion is drawn about the population is called descriptive statistics.

Inferential statistics:

A branch of statistics in which we collect, arrange and analyze the numerical data and also drawn conclusion about the population is called inferential statistics.

Applied statistics:

Applied statistics is the branch of statistics that make use of statistical methods and general rules in the investigation of a specific problem. This branch of statistics is applied in the field of agricultural, industry, transport, business, banking, insurance, economics and social science.

Population:

A set of all possible observations whether finite and infinite relevant to characteristics of interest is called population.

i. The population of stars in the sky.
ii. The number of books in the library.
iii. The number of student in the class.

Sample:

A representative of a subset of the population is called sample.

i. The few drop of blood taken from a human body.
ii. Few plants selected from a standing crop for research purpose.

Parameter:

A numerical value calculated from the population is called parameter.

Parameters are denoted by Greek letters. i.e. μ & σ^2.

Statistic:

A numerical value calculated from sample is called statistic. Statistics are denoted by English letters$\overline{X}$ & S^2.

Variable:

A characteristic that varies from one individual to another individual or object.

i.e age, height, prices

Types of variable:

i. Qualitative variable
ii. Quantities variable

Qualitative variable:

A variable which changes only in quality from one individual to another individual and cannot be measureable is called qualitative variable. Qualitative variable is also called attribute. i.e Eye, hair color, kindness, beauty

Quantitative variable:

A variable which changes only in quantity from one individual to another individual and can be measureable is called quantitative variable.

i.e. height, weight, length, price

Random variable:

A variable which is unpredictable is called random variable.

Examples:

i. Scores of batsman.
ii. Outcomes of a die.

Constant:

A characteristic which does not vary but remains fixed is called constant.

Example:

$\pi = 22/7 = 3.1415$

Types of variable:

i. Discrete variable
ii. Continuous variable

Discrete random variable:

A countable variable is called as discrete variable. Discrete variable is also called discontinuous variable. It does not take fractional values i.e. 0.333, 1.33, 25.33

Examples:

i. No of children in a family.
ii. No of students in a class.
iii. No of tosses of a coin.

Properties:

1) The values of discrete variable are countable

2) It assumes finite number of values in the given interval

3) The graph of discrete variable is discontinuous with breaks and each point touches in the form of lines at x-axis

Continuous random variable:

A measurable variable is called continuous variable. Continuous variable takes without any gap. Continuous variable takes values as fraction, decimals, integer, and whole numbers.

Examples:

i. Speed of car.
ii. Age, height, weight. Distance

Properties:

1. It assumes all possible values without jumping on the measuring scale

2. It assumes values in whole numbers and infractions measured to any numbers of decimal place

3. It assumes infinite number of values in the given interval

4. The graph of continuous variable is a continuous curve without any breaks

Level of measurement:

if a variable in mathematics and statistics is a classification that is used to describe the nature of information contained within numbers assigned to objects and therefore within a variable.

Types of measurement:

i. Nominal
ii. Ordinal
iii. Interval
iv. Ratio

Nominal:

In this type of measurement names are assign to objects as lables.

Examples:

i. Your social security number.
ii. Yes/No
iii. Male/ Female

Ordinal:

In this classification the numbers are assigned to object represent the rank order.

Example:

i. 1^{st} 2^{nd} & 3^{rd} position in a playground.
ii. Political agreement left, center, right
iii. Level of agreement No, may be, yes

Interval:

The number assign to objects have all the features of ordinal measurement and in additional equal differences b/w measurement represent equalent.

Example:

i. Measurement of sea level.
ii. Income
iii. Time of day as 12 hours clock

Ratio:

A ratio measurement is one in which the ratio b/w any two measurements is meaningful. Example:

i. No of client past 6 months.
ii. No of question ask during the lecture.
iii. No of children.

Characteristics of Statistics:

The word Statistics in plural sense has some characteristics which are discussed below.

1. Statistics are aggregate of facts.

2. Statistics are numerically expressed.

3. Statistics deals with quantitative data only but qualitative data such as honesty, beauty, poverty, liking, intelligence, etc cannot be studied directly.

4. Statistics are collected in a systematic manner.

5. Statistics are collected for a predetermined purpose.

6. Statistics must be comparable with each other.

7. It tests the laws of other sciences.

8. Statistical test are true for the long run.

9. Statistics provides only tools for analysis.

10. Statistics are estimated according to a reasonable standard of accuracy.

Importance of statistics in different fields:

Statistics has come to play an important role in every field of life.

1. It plays an important role in business

2. All types of banks make use of statistics for a number of purposes

3. The whole structure of insurance is based on statistical data

4. Statistics has an important position in almost all the natural and social sciences.

5. Statistics has prove to be of immense use in physics and chemistry

6. Astronomy is one of the oldest branches of the statistical study

7. Statistics has prove to be of immense use in meteorology because statistical principals and method are use in weather forecasting

8. Statistical data are widely use in biology for measurement of living organisms like human beings, animals and plants etc.

9. It plays an important role in economics

10. The role of statistics in the subject of education, psychology and sociology cannot be ignored.

Limitations of statistics:

Some Limitations of statistics are listed below

1. Statistical laws are true on the average.

2. Statistics deals with aggregates.

3. Statistics cannot be applied to heterogeneous data.

4. One of the greatest limitations of statistics is that it deals with the characteristics which can be numerically specified.

5. If sufficient care is not exercised in collecting, analyzing and interpreting the data, statistical result might be misleading.

6. Only a person who has an expert knowledge of statistical methods can handle

Statistical data.

Functions or Uses of Statistics:

1. Statistics simplifies complexities.

2. Statistics presents facts in a definite form.

3. Statistics simplifies comparison of data.

4. Statistics studies relationship among different facts.

5. Statistics studies change in the level of a given phenomenon.

6. Statistics aids forecasting.

7. Statistics guides the formulation of policies.

8. Statistics tests the laws of other sciences.

Presentation of data:

The raw data arranged and reduced into a form which is easy to understand, analyze and interpret is known as presentation of data.

Principal of presentation of data:

1. Data should be arranged in such a way that it will be arouse interest in reader.
2. The data should be made sufficiently concise without losing important details.
3. The data should present in simple form to enable the reader to form quick impression and to draw some conclusion directly or in directly.
4. The data should facilitate the further statistically analysis.
5. It should define the problem and suggest it solution.

Methods of presentation of data:

Four methods of presentation of data these are

i) Classification
ii) Tabulation
iii) Graphs
iv) Diagrams or Charts.

Classification:

The process of arranging data into homogenous group or classes according to some common characteristics present in the data is called classification.

For Example: The process of sorting letters in a post office, the letters are classified according to the cities and further arranged according to streets.

Bases of Classification:

There are four important bases of classification:

(1) Qualitative Base (2) Quantitative Base (3) Geographical Base (4) Chronological or Temporal Base

(1) Qualitative Base:

When the data are classified according to some quality or attributes such as sex, religion, literacy, intelligence etc…

We use the nominal and ordinal scales in qualitative basis.

(2) Quantitative Base:

When the data are classified by quantitative characteristics like heights, weights, ages, income etc…

We use the interval and ratio scales in qualitative basis.

(3) Geographical Base:

When the data are classified by geographical regions or location, like states, provinces, cities, countries etc…

(4) Chronological or Temporal Base:

When the data are classified or arranged by their time of occurrence, such as years, months, weeks, days etc…

For Example: Time series data.

Types of Classification:

(1) One -way Classification:

If we classify observed data keeping in view single characteristic, this type of classification is known as one-way classification.

For Example: The population of world may be classified by religion as Muslim, Christians etc…

(2) Two -way Classification:

If we consider two characteristics at a time in order to classify the observed data then we are doing two way classifications.

For Example: The population of world may be classified by Religion and gender.

(3) Multi -way Classification:

We may consider more than two characteristics at a time to classify given data or observed data. In this way we deal in multi-way classification.

For Example: The population of world may be classified by Religion, gender and Literacy.

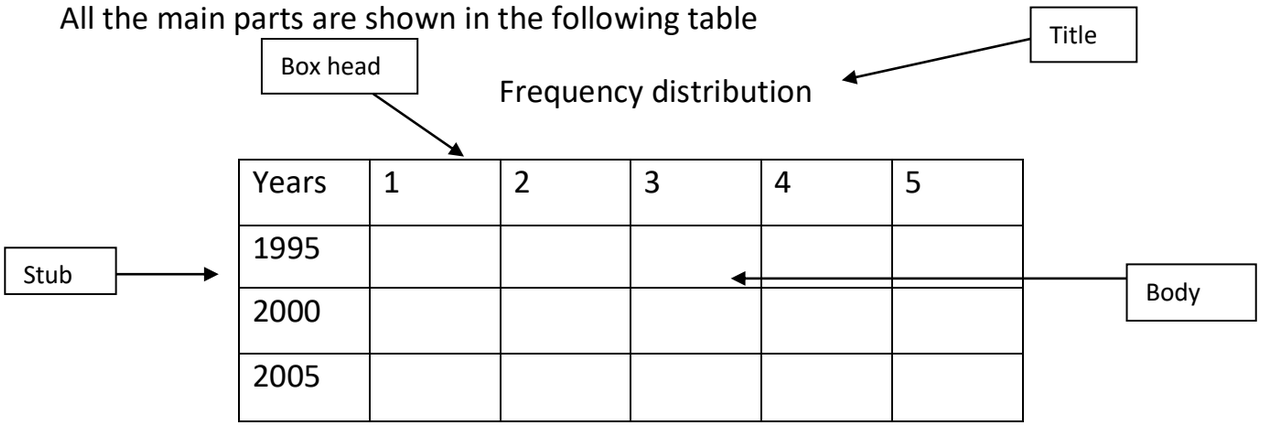
Tabulation:

The process of arranging data into rows and columns is called tabulation. A table has at least four main parts which are explained and shown in the following table.

Main parts of table:

i) Title: Heading of table written in capitals letters
ii) Stub: Row captions i.e. headings of rows.
iii) Box-head: Column captions i.e. heading of columns.
iv) Body: Main parts of table were classified data is written.
v) Source notes: It is written below the table to explain the source of data.
vi) Use of Zeros: When there is no entry in any cell of the table it should be indicated by…. Or --- but not zero (0)

All the main parts are shown in the following table



Years	1	2	3	4	5
1995					
2000					
2005					

Un-grouped data:

Initially collected data in first creative form that is not changed into any other form is called un-grouped data or raw data.

Example: 5,9,14,17,18,20

Grouped data:

Data arranged into different classes with their frequencies is called grouped data.

For example:

Classes	0-5	6-10	11-15	16-20	21-25	26-30
frequency	12	15	47	95	14	36

Array:

Data arranged in ascending or descending order of magnitude is called an array.

Example: 5, 7,9,6,4,3,1,2.

Array ascending order

1, 2,3,4,5,6,7,9

Array descending order

9, 7,6,5,4,3,2,1

Frequency distribution:

A tabular arrangement of data into classes along with corresponding frequencies is called frequency distribution.

Continuous frequency distribution:

If data of continuous variable is arranged along with corresponding frequencies is called continuous frequency distributions.

Example:

Classes	0-5	6-10	11-15	16-20	21-25	26-30
frequency	12	15	47	95	14	36

Discrete frequency distribution:

If data of discrete variable is arranged along with corresponding frequencies is called discrete frequency distributions or discontinuous distribution. Discrete frequency distribution can be converted into raw data or un-group data but continuous data cannot be converted.

Example:

Age	5	10	15	20	25	30
frequency	12	15	47	95	14	36

Example: Make an array and classify the following data into “5” classes.

4	15	16	18	27	10	18
5	7	25	14	21	30	16
2	14	23	26	29	32	17
6	9	22	24	20	13	11

Arrange the data

2	4	5	6	7	9	10
11	13	14	14	15	16	16
17	18	18	20	21	22	23
24	25	26	27	29	30	32

Now we construct the frequency distribution

i) Range = Xm-X0=32-2=30; X0= smallest value Xm= largest value

ii) Number of classes =m=1+3.3 LogN =1+3.3log (28) = 5.77

iii) Class interval =$h = \frac{range}{number\ of\ classes} = \frac{30}{5} = 6$

Marks	Tally marks	frequency
0-5		
6-10		
11-15		
16-20		
21-25		
26-30		
31-35		

Frequency:

The number of items falling in a particular class is called frequency. It is denoted by " f ".

Cass limits:

The two limits of a class are called class limits. The smaller number is called lower class limit and larger number is called upper class limit.

Class interval:

i) The difference between upper class boundary and lower class boundary of a class is called class interval. It is denoted by "h".

ii) The difference between two successive upper class limits is called class interval.

iii) The difference between two successive lower class limits is called class interval.

iv) The difference between two successive mid points is called class interval.

Note: Class interval is called length of a class.

Class marks or Mid-point:

The class marks is obtained by dividing the sum of lower limits or lower class boundary and upper class limit or upper class boundary by 2. It is denoted by "X".It is also called midpoint.

Class frequency:

The number of values falling in a specified class is called class frequency or frequency.

Class boundary:

The true class limits of a class are called class boundaries. The smaller class boundary is called lower class boundary and larger class boundary is called upper class boundary.

Example:

Marks	Class boundaries	Mid point	Tally marks	frequency
0-5	0.5-5.5	3		
6-10	5.5-10.5	8		
11-15	10.5-15.5	13		
16-20	15.5-20.5	18		
21-25	20.5-25.5	23		
26-30	25.5-30.5	28		
31-35	30.5-35.5	33		

Open end frequency distribution:

A frequency distribution that does not contain lower limit of 1st class or upper limit of last class or both is known as open end frequency distribution.

Example:

Classes	Below 5	5-9	10-14	Above 15
f	4	5	7	9

Cumulative frequency:

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

Cumulative frequency distribution:

There are two types of cumulative frequency distributions.

a) Less than cumulative frequency distributions

b) More than cumulative frequency distributions or Dissimulative cumulative frequency distributions.

Relative frequency distribution:

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

$$reletive\ frequency = \frac{frequency}{total\ frequency}$$

Relative cumulative frequency distribution:

The cumulative frequency of a class divided by total frequency is called relative Cumulative frequency and its distribution is called relative cumulative frequency Distribution.

Percentage frequency distribution:

The frequency of a class divided by the total frequency and expressing the results in percentage is called percentage frequency distribution.

$$percentage\ frequency = \frac{frequency}{total\ frequency} \times 100$$

Classes	**f**	**Cumulative freq**	**Relative freq**	**Relative cumulative freq**	**Percentage freq**
0-5	2	2	2/20	2/20	$\frac{2}{20} \times 100 = 10\%$
6-10	4	6	4/20	6/20	$\frac{4}{20} \times 100 = 20\%$
11-15	6	12	6/20	12/20	$\frac{6}{20} \times 100 = 30\%$
16-20	8	20	8/20	20/20	$\frac{8}{20} \times 100 = 40\%$

Stem and leaf display:

John Tukey 1977 introduce a technique Known as a stem and leaf display.This technique offers a quick a novel way for simultaneously sorting and displaying data set where each number in the data set is dividing into two parts a stem and a leaf. A stem is the laeding digit of each number and is used in sorting while leaf is the rest of the number and shown in display.

Example:

48,31,54,37,18,64,61,43,40,71,51,12,52,65,53,42,39,62,74,48,29,67,30,49,68,35,57,26,27,58

Arranging the data:

12,18,26,27,29,30,31,35,37,39,40,42,43,48,48,49,51,52,53,54,57,58,61,62,64,65,67,68,71,74

Stem	Leaf
1	2,8
2	6,7,9
3	0,1,5,7,9
4	0,2,3,8,8,9
5	1,2,3,4,7,8
6	1,2,4,5,7,8
7	1,4

Diagrammatic representation:

Visual display of data in the form of lines, separated bars, subdivided bars, circles or two or three dimensional geometrical forms is called chart or diagrams.

Types:

i) Linear or One dimensional diagrams

ii) Areal or two dimensional diagrams

iii) Cubic or three dimensional diagrams

iv) Pie-diagrams

v) Pictograms

vi) Cartograms

The important points for Construction all types of bar charts or diagrams

i) Title: The title should be written on the chart, so that the reader can know about the information it contains.

ii) Scale: An appropriate scale should be chosen so that the data can be represented easily on the space available on the paper.

iii) Arrangements: The arrangements of the chart should be in the most logical manner. If the data does not relate to time, it should be arranged in ascending Or descending order before charting.

iv) Source: The source of the data, for which the chart was constructed, should be indicated at the bottom of the diagram.

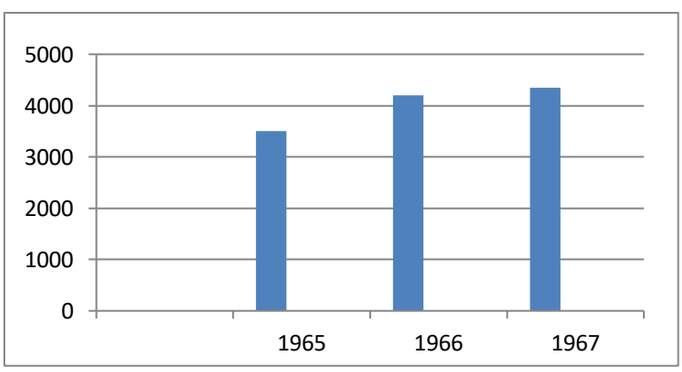
v) Mechanical details: The bars should be constructed thick lines than any other line on the chart.

Simple bar chart:

A simple bar chart consists of horizontal or vertical bar of equal width and length proportional to the values they represent. It is called simple bar chart.

Years	1965	1966	1967
Income	3500	4200	4350

Simple Bar Chart



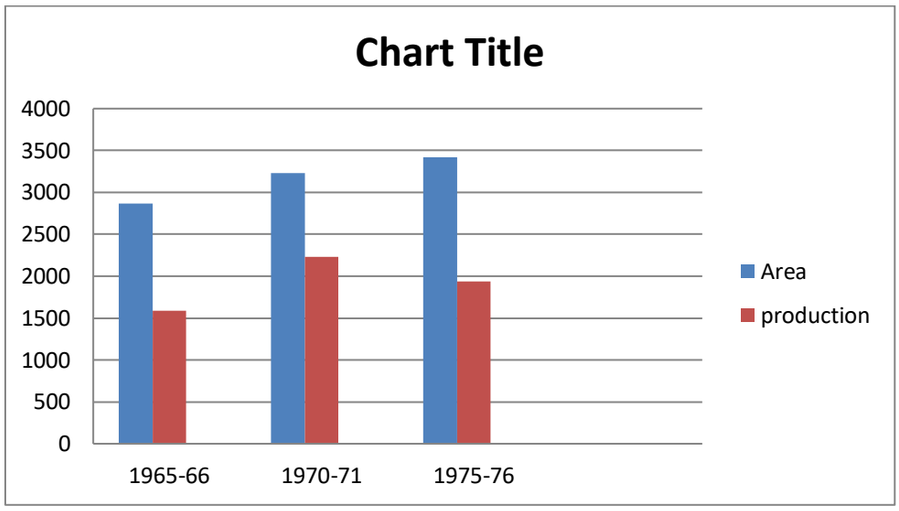
Multiple bar charts:

Multiple bar charts are merely an extension of simple bar chart. It is used to represent two or more related sets of data in the form of groups of simple bars. Each bar is shaded colored differently for identification. It is a good device for the comparative study of two or more characteristics of information.

Example: import export and production of the country can be compared from year to year by grouping the three bars together.

Year	Area	Production
1965-66	2866	1588
1970-71	3233	2229
1975-76	3420	1937

Multiple Bar Chats



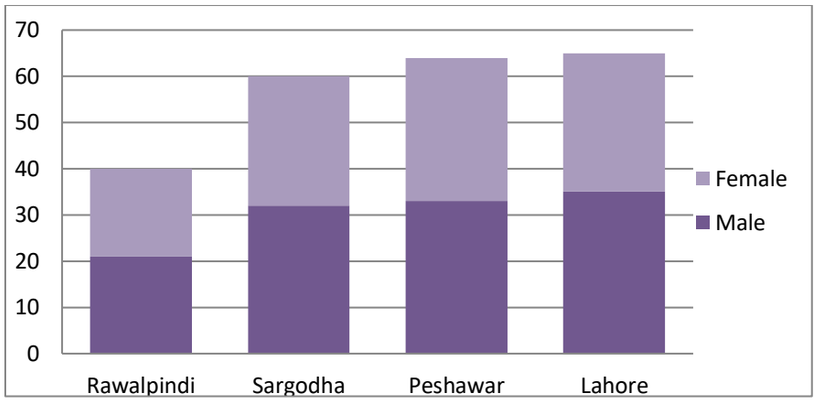
Sub-divided bar chart (component bar chart):

Sub divided bar chart is the simple bar chart which represent the total quantity of different classes and which is further sub divided into sections in the ratios of their component parts. The various component parts of the bars are shaded or colored differently for identification.

Example: The following table shows the enrollment of students by gender in institute of education and research, Punjab university, Lahore from 1975-76 to 1977-78.Draw a sub divided bar chart.

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

Component Bar Chart



Rectangles and sub-divided rectangles:

Like the component bar charts percentage sub-divided rectangles are drawn where the total values along with their components are to be compared. The comparison of budget expenditures of different families is made effectively with this chart. it is also called percentage sub divided.

The following steps involved in this construction

i) Data should be arranged in ascending or descending order

ii) Change each component into percentage of its total by using the formula

$$Percentage\ of\ component = \frac{Component\ value}{Total\ Value} \times 100$$

iii) Taking equal lengths equal to 100 units and breadths proportional to the total values draw rectangles of these breadths on the same horizontal line.

iv) Divide each of these rectangles according to the respective percentages of different component values.

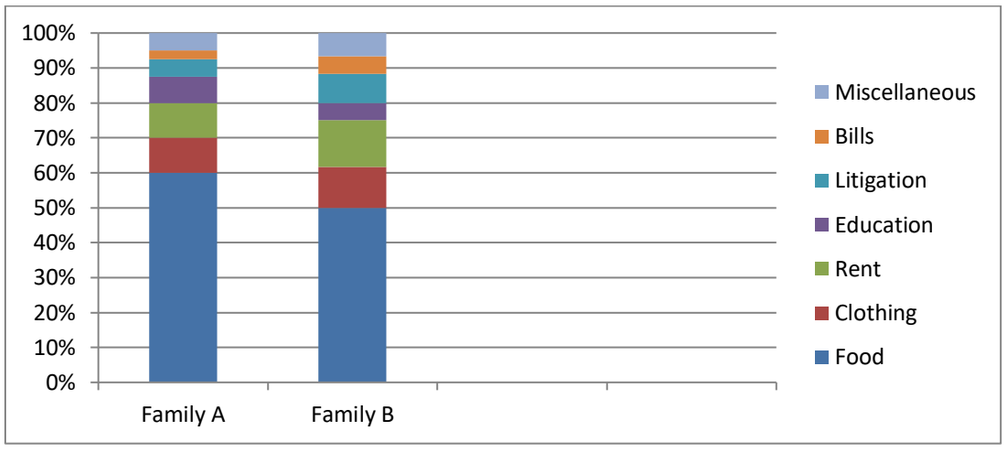
v) Each part shaded or colored will represent percentage size of one component.

Example: Compare the budgets of families A and B with a suitable diagram.

Expenditure	**Family A**	**Family B**
Food	24	60
Clothing	4	14
Rent	4	16
Education	3	6
Litigation	2	10
Bills	1	6
Miscellaneous	2	8
Total	**40**	**120**

Expenditure	Family A		Family B	
	Actual	**Percentage**	**Actual**	**Percentage**
Food	24	60.0	60	50.0
Clothing	4	10.0	14	11.7
Rent	4	10.0	16	13.3
Education	3	7.5	6	5.0
Litigation	2	5.0	10	8.3
Bills	1	2.5	6	5.0
Miscellaneous	2	5.0	8	6.7
Total	**40**	**100**	**120**	**100**

Sub-divided Rectangles



Graphs:

Presentation of statistical data by geometrical curves is called graphic representation of data.

Types:

i) Graph of time series ii) Graph of frequency distribution

The following are the important points on the construction of

graphs or line chart

i) Title: The title should be written on the graph, so that the reader can know about the information it contains. Sub-titles and foot notes should be added where necessary.

ii) Scale: An appropriate scale should be chosen for both the variables, so that the data can be displayed clearly on the graph paper available. The scale should be so fixed that the graph should exhibit clearly the variation present in the variables of the data. The scale should be written on the graph.

iii) Independent Variable : The independent variable should be indicated along x-axis while the dependent variable along y-axis

iv) The Axes: The axes should be clearly labeled, and these labels should clearly state both the variables and their units of measurement.

v) Dependent variable: For the dependent variable the scale must start at zero.

vi) Source : The source of data for which the graph is constructed should be indicated.

vii) Distinction : Graphs for two or more data must be clearly distinguished by drawing lines of different colour or shapes.

Types of Graphs:

Graph can be divided two types

a) Graph of time series or historical data

b) Graph of frequency distributions

Graph of time series or historical data:

A curve showing changes in the value of one or more items from one period of time to the next is known as graph of time series it is also called Historigram. Thus a Historigram displays the variations in the time series dealing with prices, production, imports, population, etc.

Construction of Historigram:

i) Time is taken along x-axis

ii) Values taken along y-axis

iii) Points are plotted and are then connected by means of straight line segments

Q1: The following table gives the number of cars produced in Germany

company during the years 1929-1936.Draw a suitable graph I.e. Historigram.

Years	1929	1930	1931	1932	1933	1934	1935	1936
No of cars	98	74	68	50	99	172	245	302

Graph of frequency distributions:

Curves are used for graphic representation not only of series but also for frequency distribution. Frequency distribution may be discrete or continuous. The graphic method can be used for both methods.

Types:

i) Histogram

ii) Frequency polygon

iii) Frequency curve

iv)Cumulative frequency polygon or Ogive

Advantages of graphic representation of frequency distributions:

i) It makes the unwieldy data readily intelligible

ii) It brings to light the salient features of the data at a glance

iii) It facilitates the comparison between two or more frequency distributions.

iv) It also useful in locating certain characteristics values of the distribution.

v) It main advantage is that is tells us, at a glance, to which type of theoretical

Distribution, an observed frequency distribution belongs.

Histogram:

Graphical representation of frequency distribution in the form of adjacent bars whose heights are proportional to the frequencies are plotted against class boundaries that are taken along X-axis is called histogram.

Construction:

i) Take the widths of the rectangles by marking off the class boundaries along

the x-axis

ii) Using an appropriate scale

iii) Draw a rectangle for each class whose height is proportional to its frequency.

To perform this action set up the scales as

$$Along\ y-axis\ one\ big\ sequare = \frac{Heightest\ frequencies}{No\ of\ big\ sequare\ used\ along\ y-axis}$$

Q: Draw histogram of the following frequency distribution

Classes	55-57	58-60	61-63	64-66	67-69
Frequency	4	6	14	12	4

Group	Frequencies	Class boundaries
55-57	4	54.5-57.5
58-60	6	57.5-60.5
61-63	14	60.5-63.5
64-66	12	63.5-66.5
67-69	4	66.5-69.5
Total	40	----------

Frequency polygon:

Graphical representation of frequency distribution in which frequencies are plotted against mid points that are taken along x-axis and plotted points are joined by straight lines is called frequency polygon.

Construction:

i) Take two classes one in the beginning and one at the end with zero

frequencies and calculate midpoints of all classes

ii) Mark off the midpoints of the classes along x-axis

iii) Mark the mid points along x-axis at equal distance

iv) Plot the frequencies along y-axis by taking the scale as

$$Along\ y-axis\ one\ big\ sequare = \frac{Heightest\ frequencies}{No\ of\ big\ sequare\ used\ along\ y-axis}$$

$$Along\ y-axis\ one\ small\ sequare = \frac{Value\ of\ big\ square}{10}$$

v) Plotted points are joined by straight line segments to get frequency polygon

Q: Construct a frequency polygon from the data given below by using mid

Points.

Class interval	Frequency	Mid point
0-10	0	5
10-20	5	15
20-30	9	25
30-40	12	35
40-50	20	45
50-60	16	55
60-70	0	65

Frequency Curve:

When a frequency polygon or a histogram constructed over less class interval made sufficiently small for a large number of observations is smoothed, it approaches a continuous curve called a frequency curve.

Cumulative frequency polygon or Ogive:

An Ogive or cumulative frequency curve is drawn on the basis of cumulative frequency.

Construct:

i) Compute the cumulative frequencies less than or equal to the different values

of the variable

ii) Mark off the values of the variable along x-axis

iii) Taking an appropriate scale plot the cumulative frequencies against the values

of the discrete variable

iv) Draw the horizontal straight lines through the resulting points parallel to the x-axis from each of the variable to its next value

v) Draw the vertical dashed lines to indicate the steps at respective values of

variable. It is important to note that the heights of the steps at respective values of the variable x represent the frequencies at these points

Q: Draw the cumulative frequency polygon or give from the following data.

Classes	F	C.f
0-10	15	15
10-20	17	32
20-30	19	51
30-40	25	76
40-50	16	92
50-60	15	107
60-70	13	120
70-80	10	130
80-90	5	135
90-100	2	137

Advantages of diagrams and Graphs:

i) Attractive impression

Beautifully and neatly constructed diagrams are more attractive than simple figures

ii) Easy to understand

iii) Graphs facilitate comparison between two or more than two statistical data

iv) Graphs are sometimes used to make prediction and forecasts

v) Certain partition values can also be locate graphically

vi) The help the reader in understanding the shape of the distribution of data

vii) They help in finding the trends and variations of the statistical data

viii) Economic laws are easy to understand by diagrams

ix) Diagrams and graphs have universal usefulness i.e. social scientist, mathematician etc

Disadvantages of diagrams and Graphs:

i) Diagrams and Graphs give only approximate results

ii) They are unable to represent as many sets of data as a table can do

iii) They are not sufficient information provide to detailed and accurate comparisons

iv) Graphs are less accurate

v) Diagrams and graphs take more time and labour than to construct tables

vi) A common man cannot draw diagrams and graphs but a technical hand can construct a diagram and graphs.