STATISTICAL DESCRIPTION OF DATA
(Introduction to Statistics)

Introduction:

The word "STATISTICS" has its origin from the following:

- Latin - STATUS
- German - STATISTIK
- French - STATISTIQUE
- Italian - STATISTA

Statistics in India

- Kautilya recorded birth and death in Arthashastra during Chandragupta Maurya's regime.
- Abul Fazal, during Akbar's regime, recorded agriculture in the book Ain-i-Akbari.

"STATISTICS" DEFINED

IN SINGULAR SENSE	IN PLURAL SENSE
It is defined as the scientific method of collecting, presenting, analyzing the data and drawing inference from the same.	By Statistics, we mean aggregate of facts which are known as "DATA" (Singular Datum).

Features of Statistics:

a) Statistics deals with masses and not individuals.

b) Statistics deals with quantitative data . Qualitative data are also to be expressed in quantitative terms.

c) It is aggregate of facts (plural sense).

d) It refers to scientific methods of analyzing data.(Singular Sense)

e) It is science as well as an art.

f) Data are affected by multiplicity of causes.

g) Data should be collected in a systematic manner and for a pre-determined purpose.

h) Data should be comparable.

i) All Statistics are Numerical Statements but all Numerical Statements are not statistics

APPLICATION OF STATISTICS

Statistics is used in

a) Mathematics

b) Economics

c) Accountancy

d) Auditing

e) Business and industry

f) Social Science

g) Medical Sciences & Biology

h) Different Statistical techniques used in Business, Economics and Industry.

i) Management.

LIMITATIONS OF STATISTICS

i. Statistics does not study qualitative phenomenon directly.

ii. Statistics does not study individuals.

iii. Statistical laws are not exact.

iv. Statistical data are liable to be misused.

v. Statistics results are true on the average sense only. They are not exact

FEW TERMS COMMONLY USED IN STATISTICS.

i. **Data :** It is a collection of observations, expressed in numerical figures, obtained by measuring or counting.

ii. **Population :** It is used to denote the totality of the set of objects under considering.

iii. **Sample :** A sample is a selected no. of individuals each of which is a member of the population. It is examined with a view to assessing the characteristics of the population.

iv. **Characteristic :** A quality possessed by an individual person, object or item of a population is called a characteristic e.g. Height, age, nationality, etc.

v. **Variable & Attribute :** Measurable characteristics which are expressed numerically in terms of some units are called as variables or variates e.g. age, height, income, etc. Non-measurable characteristics is a qualitative characteristic which is called as attribute e.g. sex, marital status, employment status, etc.

vi. **Continuous & Discrete Variable :** A variable which can assume for its value any real quantity within a specified interval is a continuous variable e.g height, weight etc and the variables which can assume only whole numbers are discrete variables
eg :-. number of members in the family, no of accidents etc.

CLASSWORK SECTION

Related MCQ's:

1. Which of the following statement is true?

   a) Statistics is derived from the French word "Statistik".
   b) Statistics is derived from the Italian word "Statista".
   c) Statistics is derived from the Latin word "Statistique".
   d) None of these

2. The word statistics is used in _____ senses, namely _____ and _____.

   a) two, singular, plural b) two, simple, complicated
   c) two, single, combined d) none of the above

3. The word statistics refers either _____ information or to a method of dealing with _____ information.

   a) absolute, actual b) quantitative, qualitative
   c) real, actual d) none of the above

4. Data can be obtained through a statistical _____.

   a) survey b) data c) methods d) none of the above

5. Statistics is considered with:

   a) Qualitative information b) Quantitative information
   c) Both a) and b) d) Either a) or b)

6. In the development of statistical methods, the greatest contribution is that of:

   a) Economists b) Mathematician
   c) Scientist d) Businessmen

7. Statistics is applied in

   a) Commerce & Industry b) Business Management
   c) Economics d) All of the above

8. Statistics can:

a) prove anything

b) disprove anything

c) neither prove nor disprove anything, is just a tool

d) none of the above

9. Statistics can best be considered as:

a) an art b) science

c) both art as well as science d) neither art not science

10. Which of the following would you regard as discrete variable:

a) height b) weight

c) number of persons in a family d) wages paid to workers

11. The distribution of wage is an example of the frequency distribution of

a) a discrete variable b) an attribute

c) a continuous variable d) either a) or c) above

12. An attribute is:

a) A measurable characteristics b) A quantitative characteristics

c) A qualitative characteristic d) All of the above

13. Annual income of a person is:

a) An attribute b) A continuous variable

c) A discrete variable d) Either b) or c)

14. Height of a person is:

a) An attribute b) A continuous variable

c) A discrete variable d) Either b) or c)

15. Nationality of a student is:

a) A continuous variable b) An attribute

c) A discrete variable d) None of the above

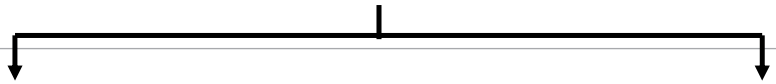
❖ **A STATISTICAL ENQUIRY PASSES THROUGH THE FOLLOWING PHASES :**

1. COLLECTION OF DATA
2. SCRUTINY OF DATA
3. CLASSIFICATION OF DATA
4. PRESENTATION OF DATA

1. COLLECTION OF DATA (DATUM IN SINGULAR)

Data : Data are aggregate of facts i.e. Quantitative information about characteristic under study.

Types of Data

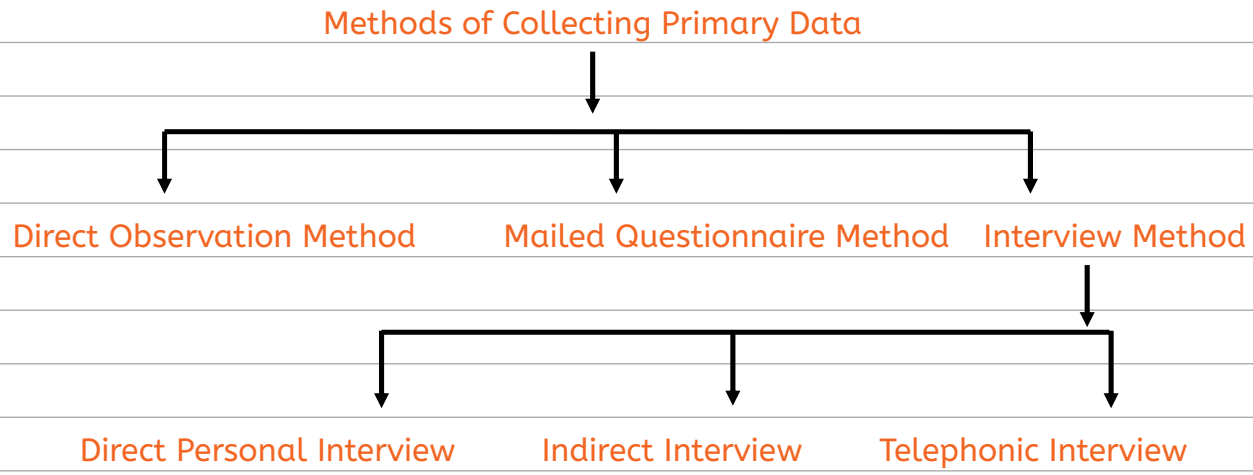


Primary Data

These data are collected for a specific purpose directly from the field of enquiry. These are original in nature

Secondary Data

1. Secondary Data are numerical information which have been previously collected as primary data by some agency for a specific purpose but are now complied from that source for use in a different connection. Sources of Secondary Data.
   - i. Publications of Central and State Governments, of Foreign Governments, and international bodies like ILO, UNO, UNESCO, WHO, etc.
   - ii. Publications of various Chambers of Commerce, Trade Associations, Co-operative Societies, etc.



(1) DIRECT OBSERVATION METHOD:

It is the best method of data collection, but time consuming, laborious and covers only a small area.

(2) MAILED QUESTIONNAIRE METHOD:

Under this method, data are collected by means of framing a well drafted and properly sequenced questionnaire covering all the important aspects of the problem under study and sending them to the respondents. (Although a wide area can be covered but non-response is maximum under this method).

(3) INTERVIEW METHOD:

a. Direct Personal Interview Method:

Under this method, the investigator collects information directly from the respondents. In case of natural calamities like earthquake, cyclone or epidemic the data can be collected much more quickly and accurately.

b. Indirect Interview Method:

It is used when the respondents can't be reached directly and the data is collected from the persons associated with the problems. E.g. in case of accidents this method is used.

Note : The above two methods are more accurate but not suitable for large area.

c. Telephonic Interview Method:

It is quick, less expensive and covers largest area. Under this method, the researcher himself gathers information by contacting the interviewee over the

phone. It is less consistent compared to the other two methods. Amount of non –response is maximum under this method.

2. SCRUTINY OF DATA

It means checking the data for accuracy & consistency. Intelligence, patience & experience is used by scrutinizing the data.

3. CLASSIFICATION OF DATA

Definitions : When the items / individuals are classified, according to some common non-measurable characteristics processed by them, they are said to form a statistical class, and when they are classified according to some common measureable characteristics processed by them, they are said to form a statistical group.

Types of Classifications

Geographical (or) Spatial i.e. Areawise	**Chronological (or) Temporal or Time Series i.e. on the basis of time**	**Qualitative (or) Ordinal**	**Quantitative(or) Cardinal**

Related MCQ's:

16. A statistical survey may either be _____ purpose or _____ purpose survey.
    a) general, specific
    b) general, without
    c) all, individual
    d) none of the above

17. Data are generally obtained from:
    a) primary sources
    b) secondary sources
    c) both primary and secondary sources
    d) neither from primary nor from secondary sources

18. Data originally collected for an investigation are known as:

a) primary data
b) secondary data
c) both primary and secondary data
d) none of the above

19. Secondary data:

a) should never be used
b) should be used after careful scrutiny
c) no scrutiny is required while using it
d) while scrutinizing the only thing to see is who collected it

20. Primary data are:

a) always more reliable compared to secondary data
b) less reliable compared to secondary data
c) depends upon the care with which data have been collected
d) depends upon the agency collecting the data

21. The quickest method to collect primary data is:

a) Personal Interview
b) Indirect Interview
c) Mailed Questionnaire Method
d) Telephonic Interview

22. In Indirect Oral Investigation:

a) Data is not capable of numerical expression
b) Not possible or desirable to approach informant directly
c) Data is collected from the books
d) None of the above

23. Some important sources of secondary data are:

a) International & Government sources
b) International and Primary sources
c) Private and Primary sources
d) Government sources

24. The data obtained by the internet are:

a) Primary data
b) Secondary data
c) Both a) and b)
d) Neither a) nor b)

25. Which method of collection of data covers the widest area?

a) Direct interview method.
b) Mailed questionnaire method.
c) Telephone interview method.
d) both (b) & (c)

26. In case of a rail accident, the appropriate method of data collection is by :

a) Direct interview
b) Personal interview
c) Indirect interview
d) All of the above

27. The best method to collect data, in case of a natural calamity, is :

a) Personal interview
b) Questionnaire method
c) Indirect interview
d) Direct observation method

28. Classification is the _____ step in tabulation.

a) first b) second
c) last d) none of the above

29. When data are observed _____ the type of classification is known as chronological classification.

a) for some hours
b) over a period of time
c) seriously
d) none of the above

30. _____ classification refers to the classification of data according to some characteristics that can be measured.

a) qualitative
b) subjective
c) quantitative
d) all of the above

31. Classification is the process of arranging data in:

a) different columns
b) different rows
c) grouping of related facts in different classes
d) different columns and rows

32. In chronological classification data are classified on the basis of:

a) attributes
b) class interval
c) locations
d) time

33. Geographical classification means classifications of data according to:

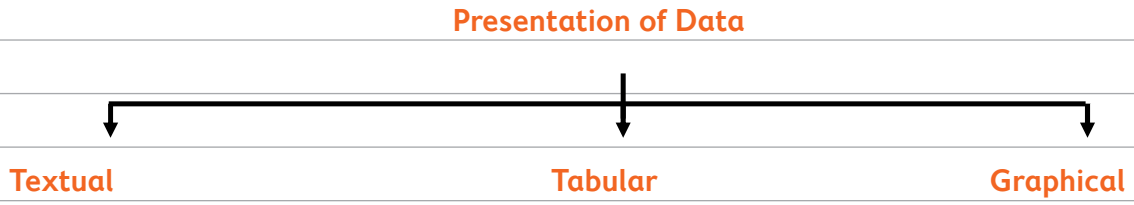
a) time
b) location
c) attributes
d) class intervals

34. The primary rules that should be observed in classification:

I. As far as possible, the class should be of equal width.
II. The classes should be exhaustive.
III. The classes should be un-ambiguously defined.

a) Only I and II
b) Only II and III
c) Only I and III
d) All I, II and III

Presentation of Data



Textual

Textual Presentation : It is in written form. It is simple but dull, monotonous & comparison is not possible

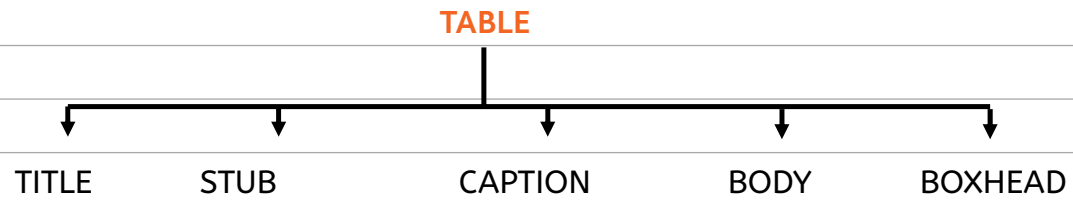
Tabular

Tabular Presentation : Presentation of data with the help of a statistical table having rows & columns.

Advantages of Tabulation are as follows:

1. Complicated data can be represented.
2. It is a must for diagrammatic representation.
3. Statistical analysis is not possible without tabulation.
4. It facilitates comparison between rows & columns.

3. DIFFERENT PARTS OF A TABLE



1. **Title :** Every Table must be given a suitable title, which usually appears at the top of the table (below the table number or next to the table number). A title is meant to describe in brief and concise form the contents of the table and should be self-explanatory.

2. **Stub :** Stubs are the headings or designations for the horizontal rows.

3. **Captions :** Captions are the headings or designations for vertical columns.

4. **Body :** The arrangement of the data according to the descriptions given in the captions (columns) and stubs(rows) forms the body of the table. It contains the numerical information which is to be presented to the readers and forms the most important part of the table.

5. **Box-head:** The entire upper part of the table is known as box-head.

Other Parts :

6. **Table Number :**

7. **Head Note :**

8. **Foot Note :**

9. **Source Note**

FORMAT OF A BLANK TABLE

Title

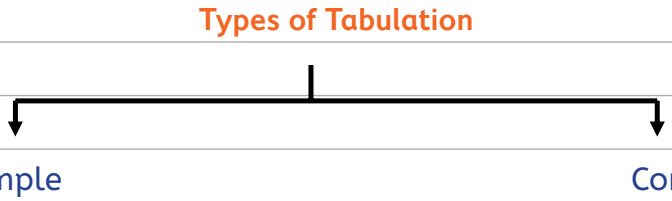
[Head Note or Prefatory Note (if any)]

Stub Heading ↓	Captions					Total
	Sub-Heads		Sub-Heads			
	Column Head	Column Head	Column Head	Column Head	Column Head	
	Body					
Total						

Foot Note :

Source Note :

Types of Tabulatio



Simple Tabulation : In this type the number or measurement of the items are placed below the headings showing the characteristics.

Complex Tabulation : In this type each numerical figure in the table is the value of the measurement having the characteristics shown both by the column and the row headings.

Related MCQ's

35. The most accurate mode of data presentation is :

a) Diagrammatic method b) Tabular

c) Textual presentation d) None of the above.

36. When the accuracy in presentation is more important than the method of presentation it is done through:

a) Textual b) Diagrammatic

c) Tabular d) Either b) or c)

37. In tabulation source of the data, if any, is shown in the :

a) Source note b) body

c) stub d) caption

38. A table is a systematical arrangement of statistical data in _____.

a) boarders and boundaries b) lanes and pillars

c) columns and rows d) all of the above

39. The unit of measurement in tabulation is shown in

a) box head b) body c) caption d) stub.

40. For tabulation, 'caption' is :

a) the lower part of the table.

b) the main part of the table.

c) the upper part of the table.

d) the upper part of a table that describes the column and sub-column.

41. The entire upper part of a table is known as :

a) caption b) stub c) box head d) body.

42. 'Stub' of a table is the

a) right part of the table describing the columns.

b) left part of the table describing the columns.

c) right part of the table describing the rows

d) left part of the table describing the rows.

43. The heading of a row in a statistical table is known as:

a) stub b) caption c) title d) foot note

44. The best method of presentation of data is :

a) Textual b) Tabular

c) Diagrammatic d) Both b) and c) above

45. In tabulation, source of data, if any, is shown in the:

a) Stub b) Body c) Caption d) Footnote

46. A table has _____ parts.

a) Two b) Three c) Four d) Five

47. The column headings of a table are known as:

a) Body b) Stub c) Box head d) Caption

Diagrammatic Representation of Data

1. Diagrammatic Representation are mainly done by charts (or graphs) and figures.

2. A chart or graph is inferior to a table or numbers as a method of presenting data, since one can get only approximate idea from it, but its advantage is that it emphasizes certain facts and relations more than numbers do.

Advantages :

1. It is more attractive and informative to an ordinary person.

2. A complex problem can sometimes be clarified easily by a diagram.

3. It reveals the hidden facts which are not apparent from the tabular presentation.

4. Two or more sets of values can be compared very easily from a diagram.

5. It shows the relation of the parts to the whole.

Types of Diagrams

Without Frequency	With Frequency (Frequency Curves)
1. Line Chart or Line Graph or Line Diagram or Historigram Chart (one dimensional)	1. Histogram or Area Diagram (Two dimensional)
2. Bar Diagram or Bar Chart (one dimensional)	2. Frequency Polygon (Two dimensional)
3. Pie Chart (Two dimensional)	3. Frequency Curve (Two dimensional)
	4. Cumulative Frequency Polygon or Ogive (Two dimensional)

Each of the Diagram is described below:

Line Diagram :

It is used for time related data (Time series).

When there is wide range of fluctuations, logarithmic or ratio charts are used.

Multiple Line Chart :

It is used for representing 2 or more related series expressed in same units.

Multiple Axis Chart :

Multiple Axis Chart is used for representing two or more related series expressed in different units.

Semi-Logarithmic Graph or Ratio Chart :

Semi-Logarithmic Graph or Ratio Chart is a line diagram drawn on a special type of graph paper which shows the natural scale in the horizontal direction and the logarithmic or ratio scale in the vertical direction. The semi-log graph is used where ratios of change are more important than absolute amounts of change.

Bar Diagram

1. **Vertical Bar Chart (or Colum Chart) :**

   This is generally used to represent a time series data or a data which is classified by the values of the variable. (Measurable characteristics).

2. **Horizontal Bar Chart :**

   This is used to represent data classified by attributes or data varying over space. (i.e. non-measurable characteristics).

3. **Grouped or Multiple or Compound Bar Chart):**

   These are used to compare related series.

4. **Component /Sub divided Bar Chart:**

   These are used for representing the data divided into different components

5. **Percentage Bars :**

   Percentage Bars are particularly useful in statistical work which requires the portrayal of relative changes.

6. **Deviation Bars**

   Deviation Bars are popularly used for representing net quantities – excess or deficit i.e. net profit, net loss, net exports or imports, etc. Such bars can have both positive and negative values. Positive values are shown above the base line and negative values below it.

7. Broken Bars

In certain series there may be wide variations in values – some value may be very small and others very large. In order to gain space for the smaller bars of the series, larger bars may be broken.

PIE CHART / PIE DIAGRAM / CIRCLED DIAGRAM

This is a very useful diagram to represent data which are divided into a number of categories. The diagram consists of a circle divided into a number of sectors whose areas are proportional to the values they represent. Again the areas of the sectors are proportional to their angles at the centre. Therefore, ultimately the angles of the different sectors are proportional to the values of different components. The total value is represented by the full circle. Comparison among the various components or between a part and the whole of data can be made easily by this diagram.

Example :

Draw a pie chart to represent the following data on the proposed outlay during a Five-year Plan of a Government :

Items	₹ (in crores)
Agriculture	12,000
Industry & Minerals	9,000
Irrigation & Power	6,000
Education	8,000
Communication	5,000

Calculations for the angles of the pie chart

Items	Outlay (in crores ₹)	Angles (in egrees)
Agriculture	12,000	108
Industry & Minerals	9,000	81
Irrigation & Power	6,000	54
Education	8,000	72
Communication	5,000	45
Total	40,000	360

Working Note : 40,000 is represented by 360^0

1,000 is represented by $\frac{360}{40} = 9^0$

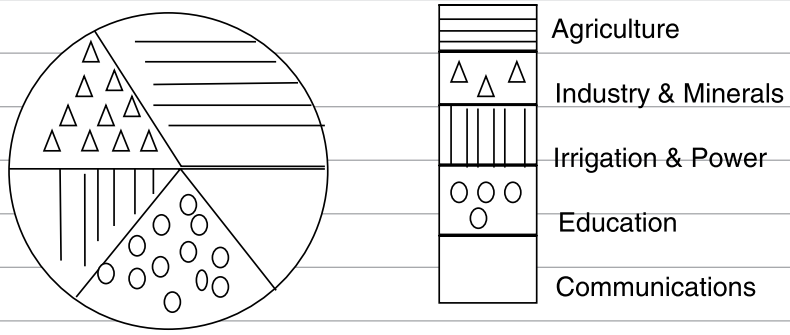
12,000 is represented by $12 \times 9 = 108^0$

9,000 is represented by $9 \times 9 = 81^0$

6,000 is represented by $6 \times 9 = 54^0$

8,000 is represented by $8 \times 9 = 72^0$

And 5,000 is represented by $5 \times 9 = 45^0$



DIAGRAMMATIC/GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTION

1. Histogram or Area Diagram

i) It consists of a set of adjoining vertical rectangles whose widths represent the class intervals and the heights represent the corresponding frequencies (for equal class width) and frequency densities (for unequal class width). Boundaries are plotted along the horizontal axis and the frequencies (or frequency densities) are plotted along the vertical axis

ii) The area of each rectangle is proportional to the frequency of the corresponding class.

iii). Mode is calculated graphically from Histogram.

iv) It helps us to get an idea about the frequency curve and frequency polygon.

v) Comparison among the frequencies can be made for different class intervals.

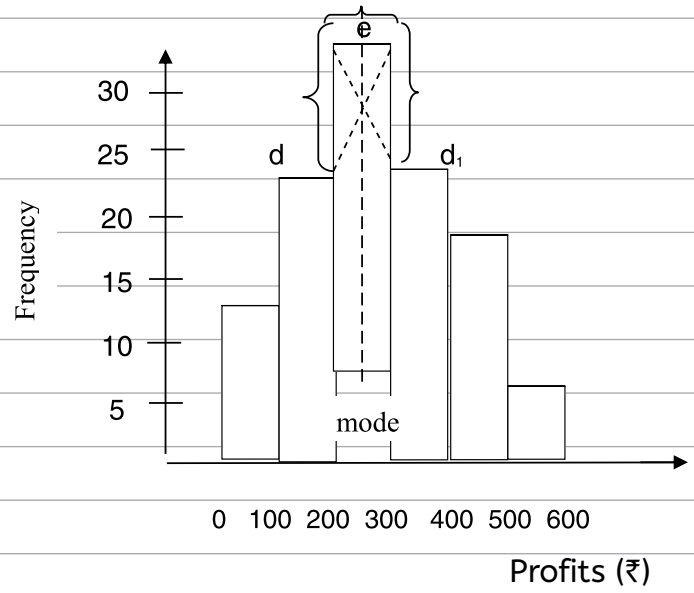
Example

The monthly profits in rupees of 100 shops are distributed as follows:

Profits per Shop	0-100	100-200	200-300	300-400	400-500	500-600
No. of Shops	12	18	27	20	17	6

Draw the histogram to the data and hence find the modal value.

In the histogram, the top right corner of the highest rectangle is joined by a straight line to the top right corner of the preceding rectangle. Similarly, top left corner of the highest rectangle is joined to the top left corner of the following rectangle. From the point of intersection of these two lines a perpendicular is drawn on the horizontal axis. The foot of the perpendicular indicates the Mode. This is read from the horizontal scale and the modal value is found to be 256 (in ₹) approximately.



2. Frequency Polygon and Frequency Curve

i) In this method, the frequency of each class is plotted against the mid-value of the corresponding class. The points thus obtained are joined successively by straight lines. The polygon is then completed by joining two end-points to the mid-values of two empty classes assumed in either side of the frequency distribution.

ii) Frequency polygon can be obtained from the histogram by joining the successive

mid-points of the top of the rectangles which constitute the histogram and the polygon is completed in the same manner as before.

iii) If in a frequency distribution the widths of the classes are reduced, then the number of classes will increase. As a result the vertices of a frequency polygon will come very close to each other. In that case, if we join the points by smooth free hand line instead of straight lines, a smooth curve is obtained which is known as a Frequency Curve.

iv) Frequency Curve is a limiting curve case of frequency polygon.

3. Cumulative Frequency Polygon / Ogive Curve

1. It is a graphical representation of cumulative frequency distribution.
2. Median and all other partition values are calculated from ogives.
3. There are two types of ogives (i) Less Than Ogive (ii) More Than Ogive.
4. IN LESS THAN OGIVE LESS THAN CUMULATIVE FREQUENCIES ARE USED. AND IN CASE OF MORE THAN OGIVE, MORE THAN CUMULATIVE FREQUENCIES ARE USED AND THE OGIVE CURVE LOOKS LIKE ELONGATED "S". THESE ARE ALSO KNOWN AS "S" CURVE.

Example

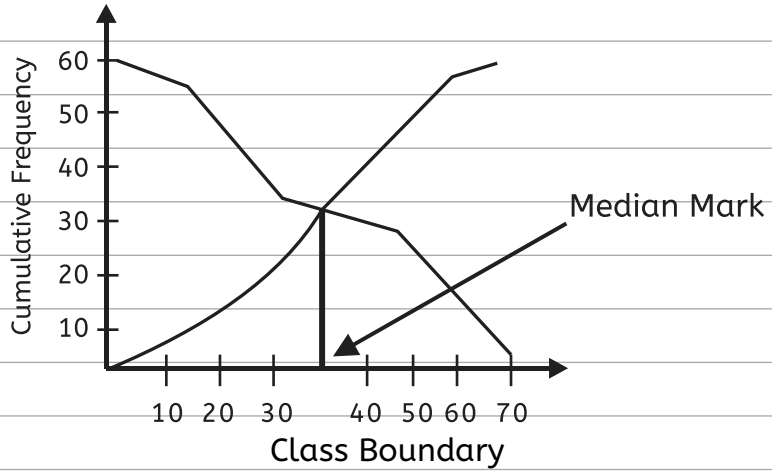
Draw the cumulative frequency diagram (both more-than and less-than ogive) of the following frequency distribution and locate graphically the Median:

Marks-Group	0-10	10-20	20-30	30-40	40-50	50-60	60-70	Total
No. of Students	4	8	11	15	12	6	3	59

Calculation for Cumulative Frequencies

Class Boundary	Cumulative Frequency	
	Less than	More than
0	0	59
10	4	55
20	12	47
30	23	36
40	38	21
50	50	9

60	56	3
70	59	0



Less than and More than ogive of a frequency distribution

From the graph the median is found to be 34.5.

4. Other Frequency Curves

1. Bell Shaped (Symmetrical Curve):

The most commonly used frequency curve use for the distribution of height, weight, profit, etc.

i. It is the limiting form of histogram and frequency polygon

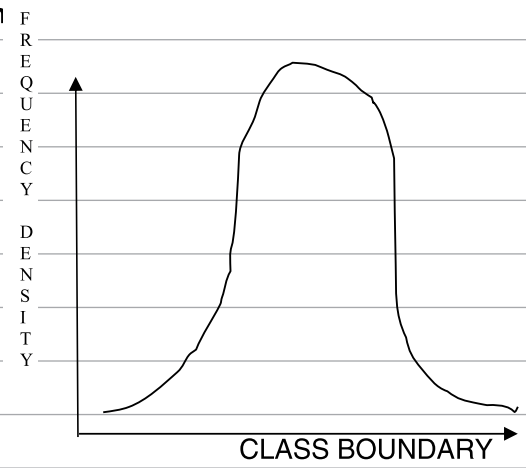
ii. The area under the curve is taken to be unity.

iii. It enables us to understand symmetry of the distribution.

TYPES OF FREQUENCY CURVES

1. Bell Shaped Cureve

Diagram

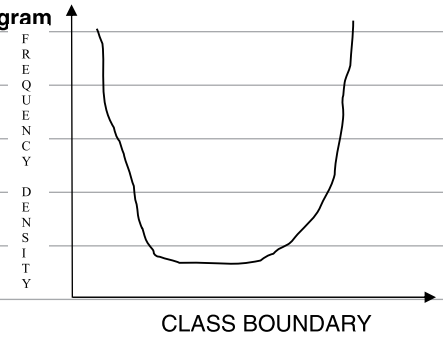


Bell Shaped Curve

2. U Shaped Curve

In this curve, the frequency is minimum at the central part, and slowly but steadily it reaches to two extremities. The distribution of people travelling on streets will be exhibited through this kind of curves.

Diagram

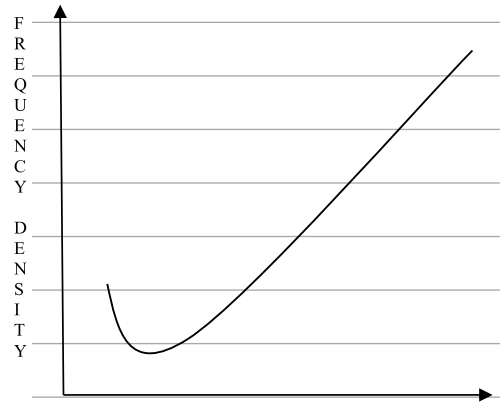


U- Shaped Curve

3. J Shaped Curve:

The J Shaped Curve starts with the minimum frequency and then gradually reaches its maximum frequency at the other extremity. The distribution of commuters in a particular time interval will be exhibited through this kind of curves.

Diagram



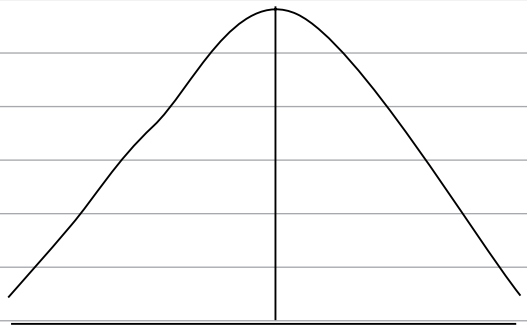
J- Shaped Curve

4. Asymmetrical Curves

(A) In case of symmetrical curves or bell shaped curves the

(i) Mean (M) = Median (Me) = Mode (Mo)

ii) Skewness = 0



$M = M_e = M_o$

Zero Skewness

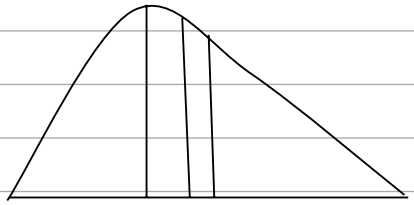
(B) In case of Asymmetrical curves Mean, Median & Mode are unequal and accordingly skewness $\neq 0$

Asymmetrical Curves

Positively Skewed

(Mean > Median > Mode)

(i) Frequency curve as a longer tail to the right



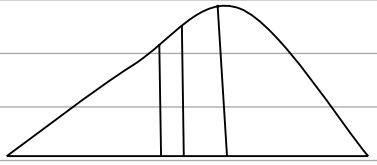
M_o M_e M

Positive Skewness

Negatively Skewed

(Mean < Median < Mode)

(i) Frequency curve as a longer tail to the Left



M M_e M_o

Negative Skewness

Related MCQ's :

48. In a two-dimensional diagram:

a) only height is considered

b) only width is considered

c) both height and width is considered

d) height, width and thickness are considered

49. Graph is a :

a) Line diagram

b) Bar diagram

c) Pie diagram

d) Pictogram.

50. The chart that uses logarithm of the variable is known as :

a) Multiple line chart
b) Ratio chart
c) Line chart
d) Component line chart.

51. Multiple axis line chart is considered when

a) the units of the variables are different.
b) there is more than one time series
c) both a) and b) above
d) either a) or b) above

52. The graphical representation of a cumulative frequency distribution is called

a) histogram
b) ogive
c) both a) and b) above
d) none of the above

53. Ogive is a

a) line diagram
b) bar diagram
c) both a) and b) above
d) none of these

54. The most common form of diagrammatic representation of a grouped frequency distribution is :

a) ogive
b) histogram
c) frequency polygon
d) none of the above

55. Frequency density is used in the construction of

a) histogram
b) frequency polygon
c) ogive
d) none of the above

56. When the width of all classes is same, frequency polygon has not the same area as the Histogram :

a) true
b) false
c) both a) and b) above
d) none of the above

57. Diagrammatic representation of the cumulative frequency distribution is :

a) frequency polygon
b) ogive
c) histogram
d) none of the above

58. A comparison among the class frequencies is possible in

a) ogive b) histogram

c) frequency polygon d) either b) or c) above

59. Mode is found graphically by :

a) frequency polygon b) ogive

c) histogram d) none of the above

60. Frequency curve is a limiting form of

a) frequency polygon b) histogram

c) either a) or b) above d) both a) and b) above

61. The breadth of the rectangle is equal to the length of the class-interval in

a) ogive b) histogram

c) both a) and b) above d) none of these.

62. Consecutive rectangles in a Histogram have no space in between them.

a) true b) false

c) both a) and b) above d) none of the above

63. Median of a distribution can be obtained from

a) histogram b) frequency polygon

c) Ogives d) none of the above

64. The curve obtained by joining the points, whose x-coordinates are the upper limits of the class-intervals and y-coordinates are corresponding cumulative frequencies is called

a) ogive b) frequency polygon

c) histogram d) frequency curve

65. The purpose served by diagrams and charts is:

a) To avoid tabulation b) To avoid textual form

c) Simple presentation of data d) All of the above

66. Less than type and more than type ogives meet at a point known as:

a) Mean b) Median c) Mode d) Quartile

67. If we plot less than and more than type frequency distribution, then the graph plotted is:

a) Frequency Curve b) Histogram

c) Ogive d) None of these

68. From which graphical representation, we can calculate partition values?

a) Lorenz Curve b) Ogive Curve

c) Histogram d) None of these

69. When the two curves of ogive intersect, the point of intersection provides:

a) First Quartile b) Third Quartile

c) Second Quartile d) None of these

70. Divided bar chart is good for:

a) Comparing various components of a variable

b) Relating the different components to the variable

c) Both a) and b) above

d) Neither a) nor b) above

71. In order to compare two or more related series, we consider:

a) Multiple Bar Chart b) Grouped Bar Chart

c) Both a) and b) d) None of them

72. An area diagram is:

a) Ogive b) Histogram c) Frequency Polygon d) None of these

73. Which of the following is a two dimensional figure?

a) Line Diagram b) Pie Diagram

c) Squares d) Both b & C

74. Arrange the dimensions of Bar Diagram, Cube Diagram, Pie Diagram in sequence.

a) 1, 3, 2 b) 2, 1, 3 c) 2, 3, 1 d) 3, 2, 1

75. The most appropriate diagram to represent the data relating to the monthly expenditure on different items by a family is:

a) Pie Diagram b) Line Diagram c) Histogramd) Frequency Polygon

FREQUENCY DISTRIBUTION

1. Tabular representation of statistical data is usually made in ascending order of magnitude relating to measurable characteristics according to individual value or group of values.

2. **There are two types of frequency distribution**
   - i. For discrete variable it is known as simple or ungrouped or discrete frequency distribution.
   - ii. For continuous random variable it is known as continuous or grouped frequency distribution.

3. **SOME IMPORTANT TERMS**

   i) **Frequency : (Tally Mark)**

   Frequency of a value of variable is the number of times it occurs in a given series of observations. A Tally Mark (/) is put against the value when it occurs in the raw data. Having occurred four times, the fifth occurrence is represented by putting a Cross Tally Mark (\ ) on the first four tally marks.

   ii) **Range :** Range of a given data is the difference between the largest measure and the smallest measure in a given set of observations.

   iii) **Class Interval (or class) :** A large number of observations having wide range, is usually classified into number of groups. Each of these groups is known as a class.

   iv) **Class frequency, Total Frequency :** The number of observations which is class contains, is known as its class frequency. The total number of observations in the frequency distribution is known as 'Total Frequency'.

   v) **Class Limit :** The two ends of a class interval are known as class limits of that class. The smaller of the two ends is called LOWER Class Limits and the greater is called Upper Class Limit. These classification are called non-overlapping or mutually inclusive classification.

   vi) **Class Boundaries :** When we consider a continuous variable, the observation are recorded nearest to a certain unit. For example, let us consider the distribution of weight of a group of persons. If we measure the weight nearest to the pound, then a class interval like (100-109) will include all the observations between

99.5 lb to 109.5 lb. Similarly, all the observations between 109.5 lb to 119.5 lb will be included in the class interval (110- 119). For the class interval (100-109), 99.5 is the lower class-boundary and 109.5 is the upper class boundary. For the class (110-119), the lower and upper class boundary respectively 109.5 and 119.5. These classifications are called overlapping or mutually exclusive classification.

Class boundaries can be calculated from the class limits by the following rule:

Lower Class boundary = Lower Class limit - $\frac{1}{2}$ d;

Upper Class boundary = Upper Class limit + $\frac{1}{2}$ d;

where, d is the common difference between the upper limit of a class and the lower limit of the next class. d/2 is called the Correction Factor

vii) **Mid-value (or class mark or mid point or class point) :**

Mid-value is the mid-Point of the class interval and is given by Class Mark= $\frac{UCL+LCL}{2}=\frac{UCB+LCB}{2}$

viii) **Width or Size :** This is the length of a class and is obtained by the difference between the upper and lower class boundaries of that class.

Class width / size = Difference between 2 successive LCL's / UCL's

= Difference between 2 successive LCB's / UCB's

= Difference between 2 successive mid values if all the class are of the same width.

= Difference between UCB and LCB

Note : Class width ≠ UCL-LCL

ix) **Frequency Density :** This is defined as the frequency per unit width of the class.

Frequency Density = $\frac{\text{Class frequency}}{\text{Class width}}$

It measures the concentration of the frequency of different classes.

x) **Relative Frequency :** This is the ratio of the class frequency to the total frequency,

i.e. Relative frequency $= \frac{\text{Class frequency}}{\text{Total Frequency}}$

- Relative Frequency of any class lies between 0 and 1

xi) **Percentage Frequency :**

$\frac{\text{Class frequency}}{\text{Total Frequency}} \times 100$ = or Relative frequency x 100

CUMULATIVE FREQUENCY DISTRIBUTION

1. There is another type of frequency distribution known as Cumulative Frequency Distribution where the frequencies are cumulated.

2. This distribution is prepared from the grouped frequency distribution by taking the end values (ie. class boundaries and not class limits)

3. Number of observation less than or equal to the class boundaries are called "Less-Than" Type Cumulative Frequency Distribution.

4. Number of observation greater than or equal to class boundaries are called " More-Than" Type Cumulative Frequency Distribution.

5. It can be made both for discrete series i.e. ungrouped data as well as for grouped data.

Example 2 :

From the following frequency distribution construct the cumulative frequency distribution:

Weights of 60 students in a class

Weights of 60 students in a class

Weight (kg)	Frequency
30-34	3
35-39	5
40-44	12
45-49	18

50-54	14
55-59	6
60-64	2
Total	60

Cumulative Frequency Distribution of weights of 60 students

Class Boundaries (Weight in kg)	Cumulative Frequency	
	Less Than	More Than
29.5	0	60
34.5	3	57
39.5	8	52
44.5	20	40
49.5	38	22
54.5	52	8
59.5	58	2
64.5	60	0

Otherwise

Cumulative Frequency Distribution of weights of 60 students

Class Boundaries (Weight in kg)	Cumulative Frequency	
	Less Than	More Than
30-34	3	60
35-39	8	57
40-44	20	52
45-49	38	40
50-54	52	22
55-59	58	8
60-64	60	2

Here the less than cumulative frequency of the second class is 8. This implies that there are 8 students whose weights are less than 39.5 kg (the upper boundary of that class). The more than cumulative frequency of the second class is 57, i.e. there are 57 students whose weights are more than 34.5 kg(the lower boundary of that class).

Note : **By Cumulative Frequency we usually mean less than type.**

Example 3 :

(a)	Marks	CF (Less than)	C.I	Frequency
	Less than 20	5	10-20	5
	Less than 30	18	20-30	13
	Less than 40	30	30-40	12
	Less than 50	35	40-50	5
			----	-----
				N= 35 = □f

(b)	Marks	C.I	CF (more than)	Frequency
	More than 20	20-30	35	17
	More than 30	30-40	18	8
	More than 40	40-50	10	7
	More than 50	50-60	3	3
			----	----
			CF	35

Related MCQ's

76. The number of observations corresponding to a particular class is known the _____ of that class.

a) frequency b) weight c) power d) both c) and a) above

77. the mid-point of a class is obtained by:

a) adding upper and lower limits

b) by dividing the difference of upper and lower limits by 2

c) by adding upper and lower limits and dividing it by 2

d) by deducting upper limit from the lower limit

78. The Frequency distribution of a continuous variable is known as :

a) grouped frequency distribution

b) simple frequency distribution

c) either a) or b) above

d) both a) and b) above

79. (Class frequency) / (Width of the class) is defined as:

a) Frequency density b) Frequency distribution

c) Both a) and b) above d) None of the above

80. Tally marks determines

a) class width b) class boundary

c) class limit d) class frequency.

81. Mutually inclusive classification is usually meant for

a) a discrete variable b) an attribute

c) a continuous variable d) none of the above

82. For determining the class frequency it is necessary that these classes are:

a) Mutually exclusive b) Not mutually exclusive

c) Independent d) None of these

83. Mutually exclusive classification usually meant for

a) an attribute b) a continuous variable

c) a discrete variable d) any of the above

84. The number of types of cumulative frequency is :

a) one b) two c) three d) four

85. The lower class boundary is :

a) an upper limit to Lower Class Limit

b) a Lower limit to Lower Class Limit

c) both a) and b) above

d) none of the above

86. Relative frequency for a particular class

a) lies between 0 and 1.

b) lies between − 1 and 0.

c) lies between 0 and 1, both inclusive.

d) lies between − 1 to 1.

87. In the construction of a frequency distribution, it is generally preferable to have classes of

a) equal width
b) unequal width
c) maximum width
d) none of these.

88. When one end of a class is not specified, the class is called.

a) closed-end class
b) open-end class
c) both a) and b) above
d) neither a) nor b) above

89. When all classes have equal width, the heights of the rectangles in Histogram will be numerically equal to the

a) class frequencies
b) class boundaries
c) both a) and b0 above
d) none of the above

90. The lower extreme point of a class is called :

a) lower class limit.
b) lower class boundary
c) both a) and b) above
d) none of the above

91. Most extreme values which would ever be included in a class interval are called:

a) Class Interval
b) Class Limits
c) Class Boundaries
d) None of the above

92. Frequency Density corresponding to a class interval is the ratio of:

a) Class Frequency to the Total Frequency
b) Class Frequency to the Class Length
c) Class Length to the Class Frequency
d) Class Frequency to the Cumulative Frequency

93. The upper class boundary is:

a) An upper limit to the upper class limit
b) A lower limit to the lower class limit
c) Both a) and b) above
d) None of the above

94. Mid values are also known as:

a) Lower limit b) Upper limit c) Class mark d) None

95. Length of a class is

a) The difference between the UCB and LCB of that class

b) The difference between the UCL and LCL of that class

c) Either a) or b)

d) Both a) and b)

96. For a particular class boundary, the less than cumulative frequency and more than cumulative frequency add up to

a) Total Frequency b) 50% of the total Frequency

c) Either a) or b) d) None

Theory Answers

1	b	21	d	41	c	61	b	81	a
2	a	22	b	42	d	62	a	82	a
3	b	23	a	43	a	63	c	83	b
4	a	24	b	44	b	64	a	84	b
5	c	25	d	45	d	65	d	85	b
6	b	26	c	46	d	66	b	86	c
7	d	27	a	47	d	67	c	87	a
8	c	28	a	48	c	68	b	88	b
9	c	29	b	49	a	69	c	89	a
10	c	30	c	50	b	70	c	90	b
11	c	31	c	51	a	71	c	91	c
12	c	32	d	52	b	72	b	92	b
13	b	33	b	53	d	73	d	93	a
14	b	34	d	54	b	74	a	94	c
15	b	35	b	55	a	75	a	95	a
16	a	36	c	56	b	76	a	96	a
17	c	37	a	57	b	77	c		
18	a	38	c	58	d	78	a		
19	b	39	a	59	c	79	a		
20	a	40	d	60	a	80	d		

Numerical Problems

In 1995, out of the 2,000 students in a college; 1,400 were for graduation and the rest of Post-Graduation (PG). Out of 1,400 Graduate students 100 were girls, in all there were 600 girls in the college. In 2000, number of graduate students increased to 1,700 out of which 250 were girls, but the number of PG students fall to 500 of which only 50 were boys. In 2005, out of 800 girls 650 were for graduation, whereas the total number of graduates was 2,200. The number of boys and girls in PG classes were equal.

97. Present the above information in tabular form.

98. Calculate the percentage increase in the number of graduate students in 2005 as compared to 1995.

99. Out of 1000 persons, 25% were industrial workers and the rest were agricultural workers. 300 persons enjoyed world cup matches on TV. 30% of the people who had not watched world cup matches were industrial workers. What is the number of agricultural workers who had enjoyed world cup matches on TV?

a) 230 b) 240 c) 250 d) 260

100. The class marks of a distribution are: 26, 31, 36, 41, 46, 51, 56, 61, 66, 71. Find the true class limits.

101. When the class intervals are 10 – 19, 20 – 29, 30 – 39, ... Upper class boundaries (UCB) and the Upper class limits (UCL) of the 2nd class interval are:

a) 29, 29 b) 20, 29 c) 29.5, 29.5 d) 29.5, 29

102. The class mark of the Class Intervals: 10 – 20, 20 – 30, 30 – 40, are:

a) 15, 25, 35 b) 14.5, 24.5, 34.5

c) 30, 50, 70 d) None of the above

103. From the following data find the number of class intervals, if class length is given as 5:

73, 72, 65, 41, 54, 80, 50, 46, 49, 53

a) 5 b) 6 c) 7 d) 8

104. The marks obtained by 30 students in a class test, out of 50 marks, according to their roll numbers are: 41, 25, 33, 12, 21, 19, 39, 19, 21, 12, 1, 19, 17, 12, 19, 17, 12, 17, 17, 41, 41, 19, 41, 33, 12, 21, 33, 5, 1, 21. If the data are arranged in the form of a frequency distribution with classes 1-10, 11-20, 21-30, 31-40, 41-50 then the frequencies of these 5 class intervals are:

a. 4, 13, 5, 4, 4 b. 3, 14, 5, 4, 4

c. 4, 13, 4, 5, 4 d. None of the above

105. The number of accidents for several days in a locality is given below:

No. of accidents	:	0	1	2	3	4	5	6
Frequency	:	15	19	22	31	9	3	2

What is the number of cases when 3 or less accident occurred?

a) 56 b) 6 c) 68 d) 87

106. The following data relate to the incomes of 86 persons :

Income (in ₹):	500-999	1000-1499	1500-1999	2000-2499
No. of persons:	15	28	36	7

What is the percentage of persons earning more than ₹ 1499?

a) 50 b) 45 c) 40 d) 60

107. Find the number of observations between 250 and 300 from the following data :

Value More than:	More than 200	More than 250	More than 300	More than 350
No. of observation:	56	38	15	0

a) 56 b) 23 c) 15 d) 8

108. Cost of sugar in a month under the heads raw materials, labour, direct production and others were 12, 20, 35 and 23 units respectively. The difference between their central angles for the largest and smallest components of the cost of sugar is (in degrees):

a) 48 b) 56 c) 72 d) 92

109. Sales of X Ltd for 4 months is given below:

Month	:	Jan	Feb	Mar	April
Sales (₹)	:	10,000	15,000	18,000	9,000

The above data represents:

a) Discrete Series
b) Continuous Series
c) Individual Series
d) None of the above

Fill in the following Frequency Distribution Table

Class Interval	Class Frequency	Class Limits	Class Boundaries	Mid Value	Width of the Class	Frequency Density	Relative Frequency	Percentage Frequency
1 – 5	9							
6 – 10	8							
11 – 15	9							
16 – 20	12							
21 – 25	31							
26 – 30	20							
31 – 35	11							
Total	100							

HOMEWORK SECTION

1. The quickest method to collect primary data is:

   (a) Personal interview (b) Indirect interview

   (c) Mailed Questionnaire Method (d) Telephonic interview

2. Which of the following statement is true?

   (a) Statistics is derived from the French word 'Statistik'

   (b) Statistic is derived from the Italian word 'Statista'.

   (c) Statistics is derived from the Latin word 'Statistique'.

   (d) None of these

3. The following data relates to the incomes of 90 persons:

Income in ₹ :	1500–1999	2000–2499	2500–2999	3000–3499
No. of Persons:	13	32	20	25

   What is the percentage of persons earning more than ₹ 2,500?

   (a) 45 (b) 50 (c) 52 (d) 55

4. In tabulation, source of data, if any, is shown in the:

   (a) Stub (b) Body (c) Caption (d) Footnote

5. Divided bar chart is good for:

   (a) Comparing various components of a variable

   (b) Relating the different components to the whole

   (c) (a) and (b)

   (d) (a) or (b)

6. Relative frequency for a particular class lies between:

   (a) 0 and 1 (b) 0 and 1, both inclusive

   (c) −1 and 0 (d) −1 and 1

7. Find the number of observations between 350 and 400 from the following data:

Value:	More than 200	More than 350	More than 400	More than 450
No. of observations:	48	25	12	0

(a) 13 (b) 15 (c) 17 (d) 19

8. When the width of all classes is same, frequency polygon has not the same area as the Histogram:

(a) False (b) True (c) Both (d) None

9. The graphical representation of a cumulative frequency distribution is called:

(a) Histogram (b) Ogive (c) Both (d) None

10. A table has _____ parts.

(a) Four (b) Two (c) Five (d) None

11. Cost of sugar in a month under the heads raw materials, labour, direct production and others were 12, 20, 35 and 23 units respectively. What is the difference between the central angles for the largest and smallest components of the cost of sugar?

(a) 72° (b) 48° (c) 56° (d) 92°

12. Frequency density corresponding to a class interval is the ratio of :

(a) Class Frequency to the Total Frequency
(b) Class Frequency to the Class Length
(c) Class Length to the Class Frequency
(d) Class Frequency to the Cumulative Frequency

13. In order to compare two or more related series, we consider:

(a) Multiple Bar Chart (b) Grouped Bar Chart
(c) (a) or (b) (d) (a) and (b)

14. An area diagram is:

(a) Histogram (b) Ogive
(c) Frequency Polygon (d) None of these

15. Most extreme values which would ever be included in a class interval are called:

(a) Class Interval (b) Class Limits

(c) Class Boundaries (d) None of these

16. In 2000, out total of 1,750 workers of factory, 1,200 were members of a trade union. The number of women employed was 200 of which 175 did not belong to a trade union. In 2004, there were 1,800 employees who belong to a trade union and 50 who did not belong to trade union. Of all the employees in 2004, 300 were women of whom only 8 did not belong to the trade union. On the basis of this information, the ratio of female members of the trade union in 2000 and 2004 is:

(a) 292 : 25 (b) 8 : 175

(c) 175 : 8 (d) 25 : 292

17. The lower class boundary is:

(a) An upper limit to Lower Class Limit (b) A lower limit to Lower Class Limit

(c) Both (a) & (b) (d) None of these

18. The distribution of profits of a company follows:

(a) J-shaped frequency curve (b) U-shaped frequency curve

(c) Bell – shaped frequency curve (d) Any of these

19. Out of 1000 persons, 25 per cent were industrial workers and the rest were agricultural workers. 300 persons enjoyed world cup matches on T.V. 30 per cent of the people who had not watched world cup matches were industrial workers. What is the number of agricultural workers who had enjoyed world cup matches on TV?

(a) 230 (b) 250 (c) 240 (d) 260

20. Median of a distribution can be obtained from:

(a) Histogram (b) Frequency Polygon

(c) Less than type Ogives (d) None of these

21. In indirect oral investigation:

(a) Data is not capable of numerical expression

(b) Not possible or desirable to approach informant directly

(c) Data is collected from the books

(d) None of these

22. Circular diagram are always:

(a) One-dimensional (b) Two-dimensional

(c) Three-dimensional (d) Cartograms

23. The column headings of a table are known as:

(a) Body (b) Stub (c) Box-head (d) Caption

24. Some important sources of secondary data are _____

(a) International and Government sources

(b) International and primary sources

(c) Private and primary sources

(d) Government sources

25. From the following data find the number class intervals if class length is given as 5.

73, 72, 65, 41, 54, 80, 50, 46, 49, 53.

(a) 6 (b) 5 (c) 7 (d) 8

26. The most appropriate diagram to represent the data relating to the monthly expenditure on different items by a family is

(a) Histogram (b) Pie-diagram

(c) Frequency polygon (d) Line graph

27. Which of the following is statistical data?

(a) Ram is 50 years old

(b) Height of Ram is 5´6´´ and of Shyam and Hari is 5´3´´ and 5´4´´ respectively

(c) Height of Ram is 5´6´´ and weight is 90 kg

(d) Sale of A was more than B and C

28. Sales of XYZ Ltd. for 4 months is:

Months	Sales
Jan.	10000
Feb.	15000
May	18000
Apr.	9000

The above data represents:

(a) Discrete (b) Continuous (c) Individual (d) None of these

29. Mid values are also called _____

(a) Lower limit (b) Upper limit (c) Class mark (d) None

30. Which of the following is not a two-dimensional figure?

(a) Line Diagram (b) Pie Diagram

(c) Square Diagram (d) Rectangle Diagram

31. Less than type and more than type gives meet at a point known as:

(a) Mean (b) Median (c) Mode (d) None

32. Arrange the dimensions of Bar diagram, Cube diagram, Pie diagram in sequence.

(a) 1, 3, 2 (b) 2, 1, 3 (c) 2, 3, 1 (d) 3, 2, 1

33. With the help of histogram one can find.

(a) Mean (b) Median (c) Mode (d) First Quartile

34. Nationality of a person is:

(a) Discrete variable (b) An attribute

(c) Continuous variable (d) None

35. If we plot less than and more than type frequency distribution, then the graph plotted is _____.

(a) Histogram (b) Frequency Curve

(c) Ogive (d) None of these

36. The primary rules that should be observed in classification.

(i) As far as possible, the class should be of equal width

(ii) The classes should be exhaustive

(iii) The classes should be unambiguously defined

Then which of the following is correct.

(a) only (i) and (ii) (b) only (ii) and (iii)

(c) only (i) and (iii) (d) all (i), (ii) and (iii)

37. Using Ogive Curve, we can determine

(a) Median (b) Quartile

(c) Both (a) and (b) (d) None

38. Mode can be obtained from

(a) Frequency polygon (b) Histogram

(c) Ogive (d) All of the above

39. The data obtained by the internet are

(a) Primary data (b) Secondary data

(c) Both (a) and (b) (d) None of these

40. The statistical measure computed from the sample observations alone have been termed as

(a) estimate (b) parameter (c) statistic (d) attribute

41. When the two curves of ogive intersect, the point of intersection provides:

(a) First Quartile (b) Second Quartile

(c) Third Quartile (d) Mode

42. The Choronological classification of data are classified on the basis of:

(a) Attributes (b) Area (c) Time (d) Class Interval

43. Arrange the following dimension wise: pie-diagram, bar-diagram and cubic diagram.

(a) 1, 2, 3 (b) 3, 1, 2 (c) 3, 2, 1 (d) 2, 1, 3

44. The frequency of class 20-30 in the following data is:

Class	0–10	0–20	0–30	0–40	0–50
Cumulative Frequency	5	13	28	34	38

(a) 5 (b) 28 (c) 15 (d) 13

45. The Graphical representation by which median is calculated is called

(a) Ogive Curve (b) Frequency Curve

(c) Line diagram (d) Histogram

46. From which graphical representation, we can calculate partition values?

(a) Lorenz curve (b) Ogive curve

(c) Histogram (d) None of the above

47. The data given below refers to the marks gained by a group of students:

Marks	Below 10	Below 20	Below 30	Below 40	Below 50
No. of Students	15	38	65	84	100

Then the no. of students getting marks more than 30 would be _____.

(a) 50 (b) 53 (c) 35 (d) 62

48. What is a exclusive series?

(a) In which both upper and lower limit are not included in class frequency.

(b) In which lower limit is not included in class frequency

(c) In which upper limit is not included in class frequency

(d) None of the above

49. A pie diagram is used to represent the following data:

Source of Income:	Customs	Excise	Income Tax	Wealth Tax
Amount in Crores	120	180	240	180

Angle in the pie diagram corresponding to income tax is.

(a) 120° (b) 240° (c) 180° (d) None

50. Difference between the maximum and minimum value of a given data is called

(a) Width (b) Size (c) Range (d) Class

51. If class interval is 10 – 14, 15 – 19, 20 – 24, then the first class is

(a) 10 – 15 (b) 9.5 – 14.5 (c) 10.5 – 15.5 (d) 9 – 15

52. Difference between the upper and lower boundary of a class is called _____.

(a) Class interval (b) Mid value

(c) Class boundary (d) Frequency

53. There were 200 employees in an office in which 150 were married. Total male employees were 160 out of which 120 were married. What was the number of female unmarried employees?

(a) 30 (b) 10 (c) 40 (d) 50

54. "The less than Ogive" is a:

(a) U-Shaped Curve (b) J-Shaped Curve

(c) S-Shaped (d) Bell Shaped Curve

55. To draw Histogram, the frequency distribution should be:

(a) Inclusive type (b) Exclusive type

(c) Inclusive and Exclusive type (d) None of these

56. The most appropriate diagram to represent the five – year plan outlay of India in different economic sectors is:

(a) Pie diagram (b) Histogram

(c) Line-Graph (d) Frequency Polygon

57. If the fluctuations in the observed value are very small as compared to the size of the item, it is presented by:

(a) Z chart (b) Ogive curve

(c) False base line (d) Control chart

58. For constructing a histogram, the class-intervals of a frequency distribution must be

(a) equal (b) unequal

(c) equal or unequal (d) none of these

59. 100 persons are classified into male / female and graduate / non-graduate classes. This data classification is:

(a) Cardinal data (b) Ordinal data

(c) Spatial Series data (d) Temporal data

60. If we draw a perpendicular on x-axis from the point of inter-section of both 'less than' and 'more than' frequency curves we will get the value of _____

(a) mode (b) median

(c) arithmetic mean (d) third quartile

61. Histogram is used for the presentation of the following type of series

(a) Time series (b) Continuous frequency distribution

(c) Discrete frequency distribution (d) Individual observation

62. Curves obtained by joining the points whose x coordinates are the upper limits of the class intervals and y coordinates are the corresponding cumulative frequencies is called

(a) Frequency Polygon (b) Frequency curve

(c) Histogram (d) Ogive

63. The number of observations between 150 and 200 based on the following data is

Value	More than 100	More than 150	More than 200	More than 250
No. of observations:	76	63	28	05

(a) 46 (b) 35 (c) 28 (d) 23

64. The number of car accidents in several days in a locality are given below:

No. of accidents:	0	1	2	3	4	5	6	7
Frequency:	12	9	11	13	8	9	6	3

What will be the number of cases when 4 or more accidents occurred?

(a) 32 (b) 41 (c) 26 (d) 18

65. The most common form of diagrammatic representation of a grouped frequency distribution is:

(a) Histogram (b) Ogive (c) Both (d) None

66. Classification is of _____ kinds.

(a) Two (b) Three (c) One (d) Four

67. The chart that uses logarithm of variable is known as:

(a) Ratio chart (b) Line chart

(c) Multiple line chart (d) Component line chart

68. Find the number of observation between 250 and 300 from the following data:

Value more than	200	250	300	500
No. of observation	56	38	15	0

(a) 38 (b) 23 (c) 15 (d) None of the above

69. Data collected on religion from the census reports are:

(a) Primary data (b) Secondary data

(c) Sample data (d) (a) or (b)

70. In collection of data which of the following are interview methods:

(a) Personal interview method (b) Telephone interview method

(c) Published data (d) (a) and (b)

71. Profits made by XYZ bank is different years refer to :

(a) An attribute (b) A discrete variable

(c) A continuous variable (d) None of these

72. Mode of presenting data

(a) Textual presentation (b) Tabulation

(c) Oral presentation (d) (a) and (b)

73. If the data represent costs spent on conducting an examination under various heads, then the most suitable diagram will be:

(a) Pie diagram (b) Frequency diagram

(c) Bar diagram (d) Multiple bar diagram

74. The point of intersection of less than ogive and greater than ogive curve gives us:

(a) Mean (b) Mode

(c) Median (d) None of the above

75. 'Stub' of a table is the

(a) Left part of the table describing the columns

(b) Right part of the table describing the columns

(c) Right part of the table describing the rows

(d) Left part of the table describing the rows

76. Frequency density is used in the construction of

(a) Histogram when the classes are of unequal width

(b) Ogive

(c) Frequency polygon

(d) None

77. Divided bar chart is considered for

(a) Comparing different components of a variable

(b) The relation of different components to the table

(c) (a) or (b)

(d) (a) and (b)

78. The following frequency distribution:

X	12	17	24	36	45
F	2	5	3	8	9

is classified as

(a) Continuous distribution (b) Discrete distribution

(c) Cumulative frequency distribution (d) None of the above

79. Histogram is useful to determine graphically the value of

(a) Arithmetic mean (b) Median

(c) Mode (d) None of the above

80. Data are said to be _____ if the investigator himself is responsible for the collection of the data.

(a) Primary data

(b) Secondary data

(c) Mixed of primary and secondary data

(d) None of the above

81. A suitable graph for representing the portioning of total into sub parts in statistics is:

(a) A Pie chart (b) A pictograph (c) An ogive (d) Histogram

82. The number of times a particular items occurs in a class interval is called its:

(a) Mean (b) Frequency

(c) Cumulative frequency (d) None of the above

83. An ogive is a graphical representation of

(a) Cumulative frequency distribution (b) A frequency distribution

(c) Ungrouped data (d) None of the above

84.

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency	4	6	20	8	3

For the class 20 – 30. Cumulative frequency is:

(a) 10 (b) 26 (c) 30 (d) 41

85. Which of the following graph is suitable for cumulative frequency distribution?

(a) Ogive (b) Histogram (c) G.M. (d) A.M.

86. Histogram can be shown as

(a) Ellipse (b) Rectangle (c) Hyperbola (d) Circle

HOMEWORK SOLUTION

1. (d) Telephonic interview method is considered as the quickest method to collect primary data as the relevant information can be gathered by the researcher himself by contacting the interviewer over the phone without any time log.

2. (b) According to the History of Statistics we can see that one school of thought is of the view that statistics is derived from the Italian word 'Statist'.

3. (b) No. of persons earning more than ₹ 2500 = 20 + 25 = 45

   ∴ The percentage of persons earning more than

   ₹ 2,500 = $\frac{45}{90} \times 100 = 50\%$

4. (d) The source of data, if any, in any kind to tabulation is shown in the footnote.

5. (c) Divided Bar Chart, also known as percentage Bar Diagrams, is good for both the things i.e. for comparing different components of a variable as well as the relating of the different components to the whole.

6. (b) Relative frequency of a class interval is defined as the ratio of the class frequency to the total frequency. Therefore, Relative frequency for a particular class lies between 0 and 1 both inclusive.

7. (a) The number of observation which are more than 350 in inclusive of those observation which are more than 400 and 450.

   ∴ Deducting those number of observations which are more than 400 and 450 from the number of observations which are 350, we will get the number of observations lying between 350 and 400.

   So, the number of observations lying between 350 and 400 = 25 − 12 − 0 = 13

8. (a) When the width of all classes is same frequency, polygon has the same area as the histogram.

9. (b) The graphical representation of a cumulative frequency distribution is called Ogive. i.e. by plotting the cumulative frequency against the respective class boundary, we get olives which can be less than type ogive are these than type olives depending upon the type of cumulative frequency distribution.

10. (c) A table has five parts namely.
    (i) Stub
    (ii) Caption
    (iii) Body
    (iv) Box head
    (v) Title

11. (d) Total components of the cost of sugar
    = (12 + 20 + 35 + 23) units
    = 90 units
    Largest component of cost of sugar
    = 35 units

    i.e. $\frac{12}{90} \times 360° = 140$
    Smallest component of cost of sugar
    = 12 units

    i.e. $\frac{12}{90} \times 360° = 48°$
    $\therefore$ Difference between the central angles for the largest and smallest components of the cost of sugar
    = 140° − 48° = 92°

12. (b) Frequency density of a class interval is defined as the ratio of the frequency of that class interval to the corresponding class length.

13. (c) Multiple Bar Chart also known as Grouped Bar Chart is one dimensional diagram in which two or more bars adjoining each other are constructed to represent the values of different variables or the values of various components of the same variable.
    Multiple Bar Chart or Grouped Bar Chart is considered to compare two or more related series.

14. (a) Histogram is a graph that represents the class frequencies in a frequency distribution by vertical adjacent rectangles. A Histogram is two-dimensional, i.e. a histogram comprises of both length as well as the width. As the Product of length and width indicates the area. Therefore Histogram is referred to as an Area Diagram. Its area represents the total frequency as distributed through the classes.

15. (c) Most extreme values which would be ever included in a class-interval are called as class boundaries, also referred to as actual class limit, are defined as the limits up to which the two limits, (actual) of each class may be extended to fill up the gap that exist between the classes.

16. (d) Title: Sex distribution of Trade Union and Non-union members.

Year	2000			2004		
Category	Male	Female	Total	Male	Female	Total
Member	1175	25	1200	1508	292	1800
Non-member	375	175	550	42	8	50
Total	1550	200	1750	1550	300	185

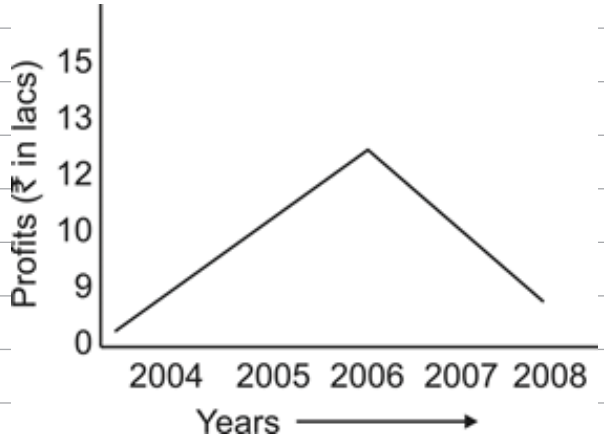
Required ratio of female members of the trade union is 2000 : 2004 = 25 : 292.

17. (b) Lower class Boundary

Lower class limit $-\frac{1}{2}$ (upper class limit to the class – lower class limit to the succeeding class). Therefore, lower class boundary is a lower limit to lower class limit.

18. (c) The bell-shaped curve looks like a bell. On a bell-shape curve, the frequency, starting from a rather low value, gradually reaches the maximum value, somewhere near the central part and then gradually decreases to reach is lowest value at the other extremity. Similar is the case of profits of a company. It rises till the resources are fully utilized and if the resources are still utilized then due to over-utilization of resources, the profits start declining. This can be clearly depicted through the data given below:

Year	Profits (` in lacs)
2004	10
2005	12
2006	15
2007	13
2008	9



19. (d)

Category	T.V.	NTV	Total
Agricultural workers	260	490	750
Industrial workers	40	210	250
Total	300	700	1000

Therefore, number of agricultural workers who had enjoyed world cup matches on T.V. = 260.

20. (c) Ogives are considered for obtaining quartiles graphically. If a perpendicular is drawn from the point of intersection of the two o-gives, i.e. less than type ogive and more than type give, on the horizontal axis, then x-value of this point gives us the value of median, the second or middle quartile.

21. (b) Indirect oral investigation is a method in which a third person is contacted who is expected to know the necessary details about the persons for whom the enquiry is meant. This method is suitable when it is not possible or deliverable to approach informant directly.

22. (b) Circular diagram is a Two-dimensions diagram in which a circle is prepared and the radius of circle is determined on the basis of minimum square root value of the variable. Two-dimensional diagram is a diagram which is prepared on the basis of two dimension i.e. length and width.

23. (d) Each column is given a heading to explain what the figures in the columns represent. These column headings of a table are known as caption.

24. (a) The Government source like Indian trade journal – weekly, reserve Bank of Indian Bulletin – monthly, etc and International sources like WHO, World Bank, IMF, etc are some of the important sources of secondary data.

25. (d) We have, Range = Maximum value – Minimum value = 80 – 41 = 39

Class length = 5

No. of class Intervals × class lengths Range

$\Rightarrow$ No. of class Intervals × 5 $\cong$ 39

$\Rightarrow$ No. of class Intervals = $\frac{39}{5}$

(We always take the next integer as the no. of class intervals so as to include both the minimum and maximum values).

26. (b) Pie diagram

27. (b) Option (b) represents statistical data which can be understood by referring the definition of statistics keeping note of the following points.

1. Statistics are aggregate of facts. A single figure cannot be called as statistics because it cannot be compared to draw any conclusion out of it.
2. All statistical facts are expressed in numbers. Qualitative expressions like young, old, etc do not constitute statistics.
3. Statistics should be placed in relation to each other so as to facilitate comparison. For this purpose, the data must be homogenous and not heterogenous. e.g. height and weight are heterogenous in character.

28. (c) Given data represents unclassified and ungrouped data. Therefore, the given series is an individual series.

29. (c) Mid-values are also called class mark.

Class Mark = $\frac{\text{Lower class limit + Upper class limit}}{2}$

30. (a) Line Diagram.

31. (b) By plotting cumulative frequency against the respective class boundary, we get Ogives. There are two type of ogives:

(i) Less than type ogive.

(ii) More than type ogive.

Olives may be considered for obtaining quartiles graphically. If a perpendicular is drawn from the point of intersection of two ogives on the horizontal axis, then then the x-value of this point gives us the value of median, the second or middle quartile.

Therefore, the meeting point of less than type ogive and more than type ogive is known as 'Median'.

32. (a) Bar diagram is one dimensional.

Cube diagram has 3 dimensions viz. length, breadth and height and hence is three-dimensional.

Pie-diagram is two-dimensional.

Therefore, if we arrange it in sequence, we get:

Bar diagram, cube-diagram and Pie diagram i.e. 1, 3, 2.

33. (c) Histogram is used to find Mode. [Self Explanatory]

34. (b) A qualitative characteristic is known as an attribute.

So the nationality of a person is an attribute as it is a qualitative characteristic.

35. (c) If we plot less than and more than type frequency distribution, then the graph plotted is Ogive.

Ogive are of two types – Less than type ogive and more than type ogive. [self-explanatory]

36. (d) Requisites of a good classification are:

1. It should be exhaustive
2. It should be mutually exclusive
3. It should be unambiguous
4. It should be stable and flexible
5. It should be homogeneous
6. It should be a revealing classification

37. (c) Olives are considered for obtaining quartiles graphically. If a perpendicular is drawn from point of intersection of two Olives on horizontal axis, then x-value of this point gives us the value of median (2nd or middle quartile).

38. (b) Mode can be obtained from histogram.

39. (b) Secondary data

40. (c) Statistic

41. (b) We know, that the two curves viz. Less than Ogive & More than Ogive intersect at a point called Median or we can say Second Quartile.

42. (c) Chronological Classification data are classified on the basis of 'Time'.

43. (d) Pie-Diagram: Two Dimensional Diagram (2)

These Diagrams are also called as "Area-Diagrams".

Used when different segments or components of values are also to be presented.

Bar-Diagram: One Dimensional Diagram (1) means such diagrams where only one dimensional measurement i.e. height is used. There is no importance of width or thickness in these diagrams. The heights of bars are taken on the basis of values.

Cubic-Diagram: Three Dimensional Diagram (3) are those in which three dimensions viz length, breadth & height are taken into account used when these is wide range of data and three different but inter-related features of data are to be represented simultaneously.

44. (c)

Class	Cumulative Freq.	Frequency
0–10	5	5
10–20	13	13 – 5 = 8
20–30	28	28 –13 = 15
30–40	34	34 – 28 = 6
40–50	38	38 – 34 = 4

45. (a) The median is calculated by Ogive Curve

46. (b) We can calculate partition values with the help of O'Give Curve for graphical representation.

47. (c) Converting the given "Less than" type frequency distribution to Normal frequency distribution:

Class Interval	(f) frequency
0 – 10	15
10 – 20	23
20 – 30	27
30 – 40	19
40 – 50	16

Hence,

The no. of students getting marks more than 30 is 19 + 16 = 35.

48. (c) In exclusive series, upper limit is not included in class frequency.

49. $\text{Angle} = \frac{\text{Revenue of Income tax}}{\text{Total Revenue}} \times 360$

$= \frac{240}{120 + 180 + 240 + 180} \times 360 \quad = \quad \frac{240}{720} \quad X\ 360 = 120$

50. (c) Difference between the maximum and minimum value of given data is called Range.

51. (b) Class intervals is 10 – 14, 15 – 19, 20 – 24

D = 15 – 14 = 1

$\frac{D}{2} = \frac{1}{2} = 0.5$

First class is (10 – 0.5) – (14 + 0.5)

= 9.5 – 14.5

52. (a) The difference between the upper and lower boundary of class is called class interval (class width).

53. (b) Total Employees in the office = 200

No. of Employees who are married = 150

No. of Employees who are unmarried = 200 – 150 = 50

No. of Total male Employees = 160

No. of Married male Employees = 120

No. of unmarried male Employees = 160 – 120 = 40

No. of females who are unmarried = 50 – 40 = 10

54. (c) "The less than Ogive" is a s-shaped.

55. (b) To Draw Histogram, the frequency distribution should be exclusive type.

56. (a) Pie diagram

57. (c) If the fluctuations in the observed value are very small as compared to the size of the item, it is present by false base line.

58. (a) For constructing a histogram, the class-intervals of a frequency distribution must be equal.

59. (b) Original data

60. (b) If we draw a perpendicular on x-axis from the point of intersection of both 'less than' and 'more than' frequency curve. We will get the value of 'Median'.

61. (b) Histogram is used for the presentation to the continuous frequency distribution of the series.

62. (d) Curve obtained by joining the points whose x co-ordinate are the upper limits of the class intervals and y co-ordinates are the corresponding cumulative frequencies is called 'o' give.

63. (b)

C.I.	Frequency
100 – 150	76 – 63 = 13
150 – 200	63 – 28 = 35
200 – 250	28 – 05 = 23
250 – 300	05

The no. of observation b/w 150 and 200 is 35.

64. (c)

No. of Accident	0	1	2	3	4	5	6	7
Frequency	12	9	11	13	8	9	6	3

No. of Cases when 4 or more Accidents occurred

= 8 + 9 + 6 + 3 = 26

65. (a) The most common form of diagrammatic representation of a group frequency distribution is Histogram.

66. (d) Classification is of four kind.

67. (a) The chart that uses logarithm of variable is known as Ratio Chart.

68. (b)

C.I.	Frequency
200 – 250	56 – 38 = 18
250 – 300	38 – 15 = 23
300 – 350	15 – 0 = 15
350 – 400	0 – 0 = 0

No. of observation b/w 250 and 350 = 23.

69. (b) Data collected on religion from the census reports are secondary data.

70. (d) Personal interview method and telephone interview method are the interview method.

71. (c) Profit made by XYZ Bank is different years refer to a continuous variable.

72. (d) Mode of presentation data are textual presentation and tabulation.

73. (a) If the data represent cost spent on conducting an examination under various heads then the most suitable diagram will be Pie diagram.

74. (c) The point of intersection of less than 0give and greater than 0give curve gives us Median.

75. (d) 'Stub' of a table is the left part of the table describing the rows.

76. (a) Frequency density is used in the construction of Histogram.

77. (d) Divided Bar Chart is considered for comparing different components of a variable and the relation of different components to the table.

78. (b) Discrete distribution

79. (c) Histogram is useful to determine graphically the value of 'mode'.

80. (a) Data are said to be Primary data if the Investigator himself is responsible for the collection of the data.

81. (a) A suitable graph for representating the portioning of total into sub parts in statistics is a Pie chart.

82. (b) The number of times a particular items occurs in a Class Interval is called its Frequency.

83. (a) An Ogive is a graphical representation of cumulative frequency distribution.

84. (c)

C.I.	F	C.F.
0 – 10	4	4
10 – 20	6	10
20 – 30	20	30
30 – 40	8	38
40 – 50	3	

Cumulative frequency of Class Interval '20 – 30' is 30.

85. (a) Ogive is graph suitable for cumulative frequency distribution.

86. (b) Histogram can be shown as Rectangle.