J.K. SHAH®
CLASSES
a Veranda Enterprise

Foundation → Intermediate → Final CA ↗

CA FOUNDATION

STATISTICS

Head Office

Shraddha, 4th Floor, Old Nagardas Road,
Near Chinai College, Andheri (E), Mumbai - 400 069.

+91 - 73044 54689

/officialjksc jkshahclasses.com

J.K. SHAH CLASSES
a Veranda Enterprise

INDEX

1 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 :

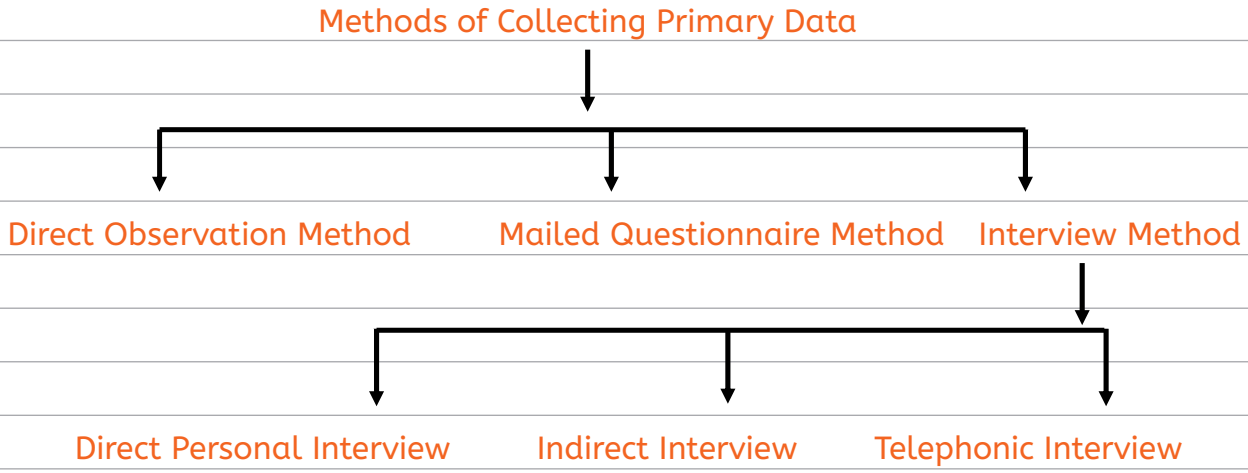
1. COLLECTION OF DATA
2. SCRUTINY OF DATA
3. CLASSIFICATION OF DATA
4. PRESENTATION OF DATA

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

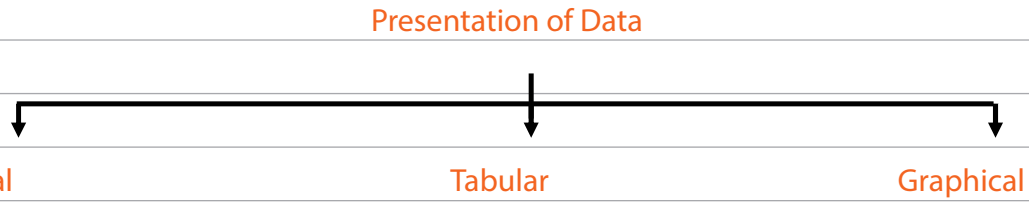
30. _____ classification refers to the classification of data according to some characteristics that can be measured.
    a) qualitative
    b) subjective
    c) quantitative
    d) all of the above

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



Textual

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

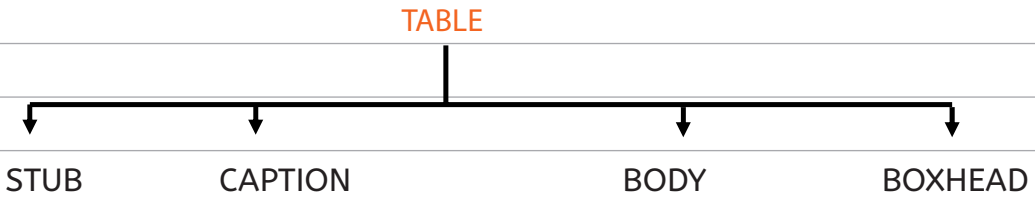
Tabular

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

2. Captions : Captions are the headings or designations for vertical columns.

3. 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.

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

Other Parts :

5. 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.

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)]

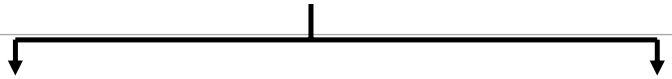
Stub Heading ↓	Captions					Total
	Sub-Heads		Sub-Heads			
	Column Head	Column Head	Column Head	Column Head	Column Head	
	Body					
Total						

Foot Note :

Source Note :

Types of Tabulatio

Types of Tabulation



Simple Complex

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 $\frac{360}{40} = 9^0$

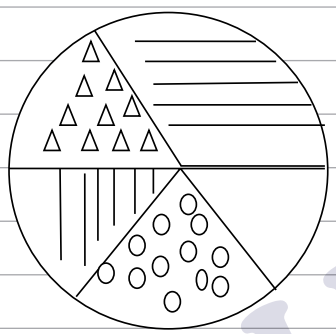
12,000 is represented by $12 \times 9 = 108^0$

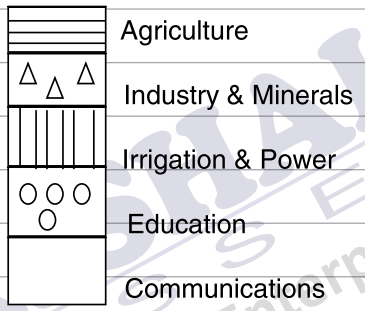
9,000 is represented by $9 \times 9 = 81^0$

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.

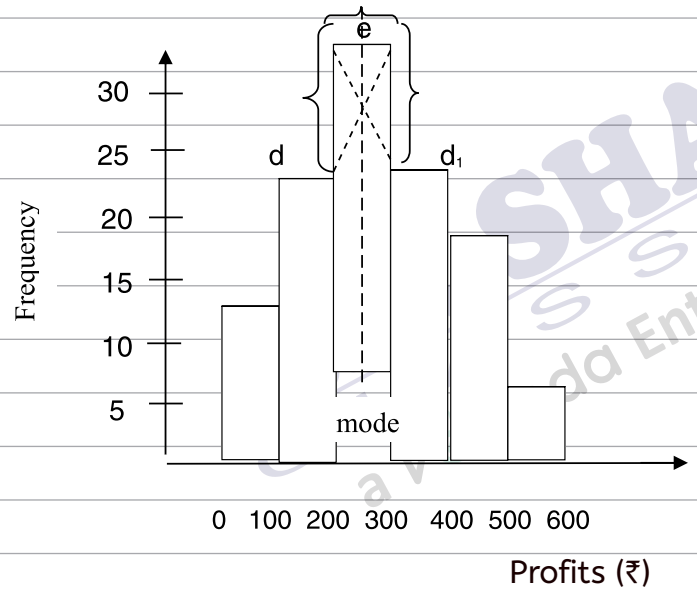
Example

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.

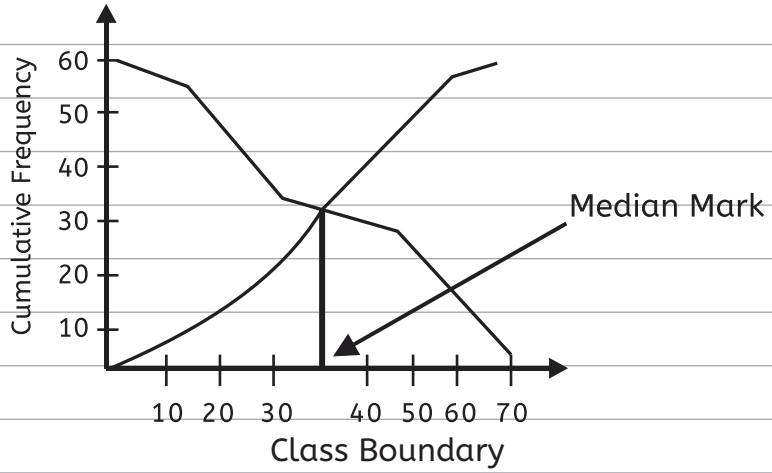
Example

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):

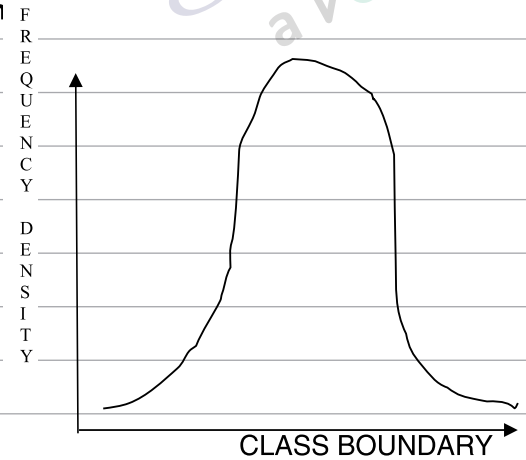
The most commonly used frequency curve use for the distribution of height, weight, profit, etc.

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.

Diagram

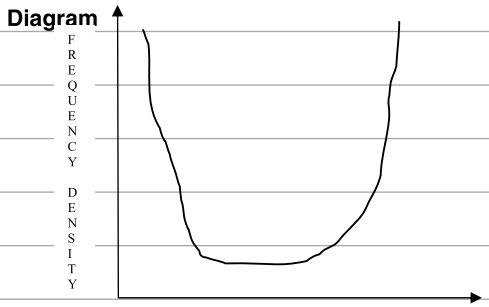


Bell Shaped Curve

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

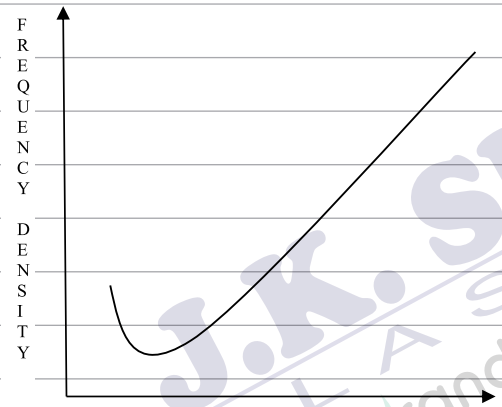


U- Shaped Curve

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



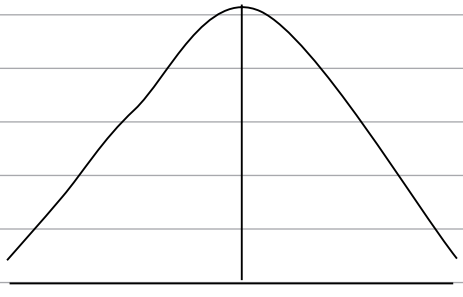
J- Shaped Curve

4. Asymmetrical Curves

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

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

ii) Skewness = 0

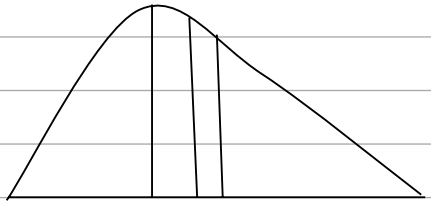


Zero Skewness

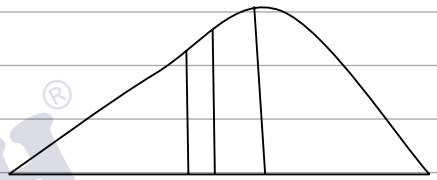
(B) In case of Asymmetrical curves Mean, Median & Mode are unequal and accordingly skewness $\neq 0$

Asymmetrical Curves

Positively Skewed	Negatively Skewed
(Mean > Median > Mode)	(Mean < Median < Mode)
(i) Frequency curve as a longer tail to the right	(i) Frequency curve as a longer tail to the Left



M_o M_e M
Positive Skewness



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) Histogram d) 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		

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 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 four parts namely.
    (i) Stub
    (ii) Caption
    (iii) Body
    (iv) Box head

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. $\frac{12}{90} \times 360° = 48°$
    ∴ 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

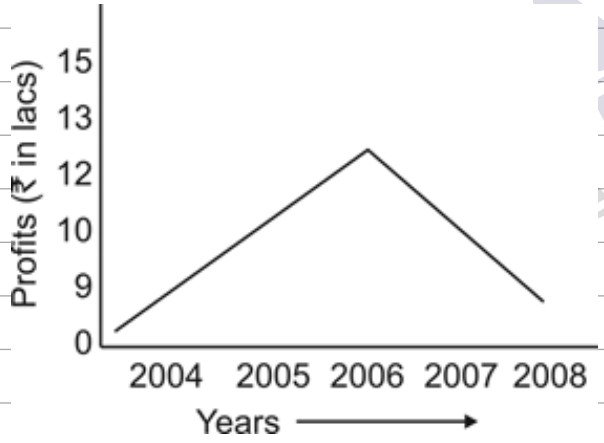
Required ratio of female members of the trade union is 2000 : 2004 = 25 : 292.

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 $\geq$ 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.

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

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

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

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.

SELF ASSESSMENT TEST 1
25 Marks

1. The meaning of STATISTICS in plural sense is:
   a) A set of numerical figures, related to any country
   b) A set of artificial figures, related to any city
   c) A set of numerical figures, related to any sphere of enquiry
   d) A set of artificial figures, related to any sphere of enquiry

2. Which of the following best describes a frequency table of a variable?
   a) A two-way classification table.
   b) A table of frequency.
   c) A one way classification table based on a variable, classified into class intervals with the cor-responding class frequency.
   d) A two-way classification table with frequencies.

3. The reason of mistakes in collection of primary data is because of:
   a) Carelessness of investigators
   b) Carelessness of informants
   c) Biasness of investigators
   d) All of the above

4. The method of presenting the classified data is:
   a) Tabulation
   b) Diagrammatic presentation
   c) Graphic presentation
   d) All of the above

5. If information is to be collected from educated people in a large area, suitable method shall be:
   a) Census
   b) Direct personal investigation
   c) Questionnaire
   d) Through correspondents

6. As the numbers of observations and classes increase, the shape of a frequency polygon:
   a) tends to become increasingly smooth
   b) tends to become jagged
   c) stays the same
   d) varies only if data become more reliable

7. Which one of the following statement is not CORRECT?
   a) Indirect oral investigation brings in bias of investigator
   b) Questionnaires through investigators is a suitable method in case of extensive enquiries
   c) Pre-testing of questionnaires is essential for a good questionnaires
   d) None of the above

8. Basis of classification of data under chronological classification shall be:
   a) According to place
   b) According to quality
   c) Magnitude of classes
   d) None of the above

9. Publication of data by the Department of Economics and Statistics are called:
   a) Departmental data
   b) Internal data
   c) Secondary data
   d) Primary data

10. Charts and graphs are the presentation of numerical facts by means of:
   a) Symbols
   b) Points and Lines
   c) Area and other Geometrical forms
   d) All of the above

11. The diagrams for the preparation of which both length and width are considered, are called:
   a) Sub-divided bar diagram
   b) Multi bar diagram
   c) Percentage sub-divided diagram
   d) Two-dimensional diagram

12. Continuous data are differentiated from discrete data in that:
   a) Discrete data can take on any real number
   b) Continuous data take on only whole numbers
   c) Continuous data classes may be represented by fractions
   d) Discrete data classes are represented by fractions

13. Class width is measured as:
   a) Half of the sum of lower and upper limits
   b) Half of the difference between upper and lower limits
   c) The difference between upper and lower boundaries
   d) The sum of the upper and lower limits

14. Class-mark is obtained:

a) By multiplying the upper and lower limits

b) By deducting lower limit from the upper limit

c) By dividing the total of upper and lower limit by 2

d) By dividing the difference of upper and lower limit by 2

15. When there is an increase in a series at constant rate, the graph will be:

a) Convex curve b) Concave curve

c) A straight line from left bottom to right top

d) A straight line from left top to right bottom

16. With the help of histogram we can prepare:

a) Frequency polygon b) Frequency curve

c) Frequency distribution d) All of the above

17. Non-dimensional diagrams are also known as:

a) Cubes b) Spheres c) Pictograms d) All of the above

18. The headings of the rows given in the first column of a table are called:

a) Captions b) Sub-titles c) Stubs d) Prefatory notes

19. Ogives for more than type and less than type distribution intersect at:

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

20. The following frequency distribution is classified as:

X:	0	1	2	3	4	5	6
F:	2	5	8	10	19	16	13

a) Cumulative frequency distribution b) Continuous distribution

c) Discrete distribution d) Median

21. The following series is of the type of:

Year	Population in city A
2010	29,00,000
2011	31,25,000
2012	35,00,000
2013	37,20,000

a) Individual Series b) Discrete Series

c) Geographical Series d) Time Series

22. The following distribution is of which type:

Vales (less than)	20	25	30	35	40	45	50
Frequency	5	10	15	20	25	30	45

a) Discrete type b) Inclusive class type

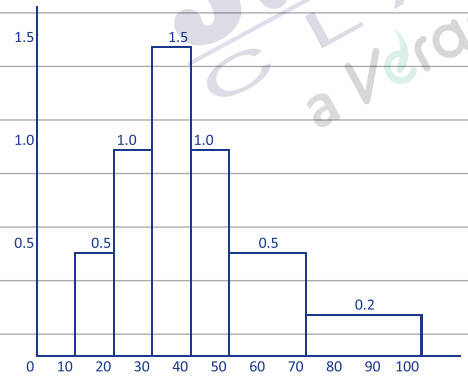
c) Exclusive class type d) None of the above

23. Find the value of F from the following distribution:

X	Frequency	Cumulative Frequency
3	5	5
6	7	12
8	6	F
12	3	21

a) 12 b) 18 c) 21 d) 24

24. The given histogram shows a frequency distribution of marks obtained by 56 students in statistics. Find the number of students securing marks between 70 and 100.



a) 2 b) 4 c) 6 d) 8

25. Study the following data. If data is presented on a pie-chart. Find the difference between central angle of C and E.

City	A	B	C	D	E
Population (in Cr)	10	6	1	0.5	0.5

a) 20° b) 10° c) 120° d) 100°

EXPLANATORY ANSWERS

1. In plural sense, statistics means a set of numerical figures, related to any sphere of enquiry. Option C

2. A frequency table of a variable is a one way classification table based on a variable, classified into class intervals with the corresponding class frequency. Option C

3. Mistakes in collection of primary data can be due to error by the investigators and informants both. Option D

4. Classified data can be presented by means of any graphs or tables or diagrams. Option D

5. Questionnaire is the best method to collected data from large educated population with least cost and in minimum time. Option C

6. As the number of observations and class increase, the frequency polygon tends to becomes flat or starts smoothening. Option A

7. All the three statements are correct. None of them is NOT CORRECT, Option D

8. Chronological classification is based on time frame or time series. Option D

9. Publication of data by the Department of Economics and Statistics are secondary in nature. Option C

10. Both charts and graphs use symbols, points, lines, area, and geometric forms. Option D

11. Two-dimensional diagrams use both length and width of the data. Option D

12. In continuous data classes may be represented by fractions but not in discrete data classes. Option C

13. Class width or length of a class interval is the difference between upper and lower boundaries of the class. Option C

14. Class-mark or mid-point of a class is obtained by dividing the total of upper and lower limit by 2. Option C

15. An increasing graph at a constant rate is shown by a upward sloping straight line from left to right on a graph paper. Option C

16. With the help of histograms, one can draw a frequency curve, frequency polygon and even a frequency distribution. Option D

17. Pictograms are the non-dimensional diagrams. Option C

18. The heading of the rows given in the first column are mentioned in stub of a table. Option C

19. The intersecting point of the more than type and less than type Ogives gives Median of the distribution. Option D

20. The given distribution is for discrete data. Option C

21. The given series is time based. It is a time series data. Option D

22. The given distribution is a cumulative frequency distribution for exclusive type class. Option C

23. F = 12 + 6 = 18. Option B

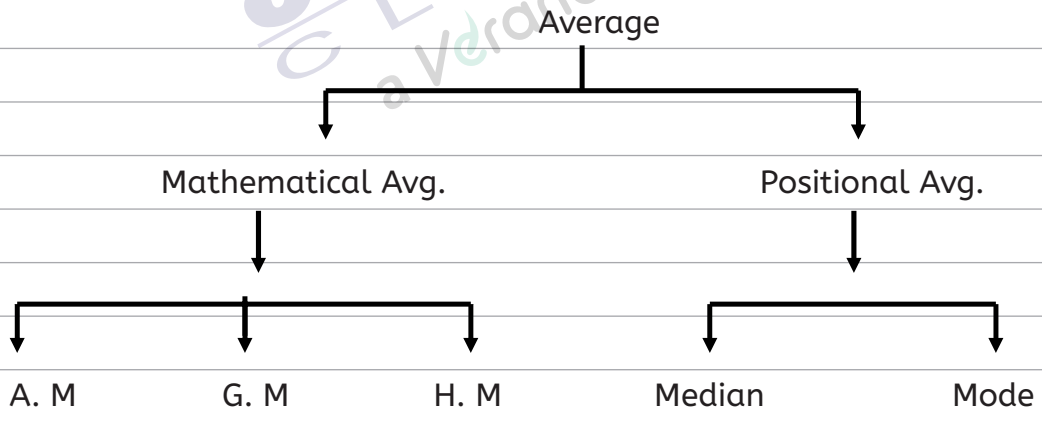
24. Total frequency = 56. Total frequency on graph = 0.5 + 1 + 1.5 + 1 + 0.5 + 0.5 + 0.2 + 0.2 + 0.2 = 5.6. Thus 0.1 points on graph represents 1 unit of frequency. Students securing marks between 70 and 100 = 0.2 + 0.2 + 0.2 = 0.6 points of graph = 6 students. Option C

25. Total population = 10 + 6 + 1 + 0.5 + 0.5 = 18 representing 360° on the pie chart. Difference be-tween C and E = 1 – 0.5 = 0.5 Cr = 0.5 * 360°/18 = 10° Option B

2 MEASURES OF CENTRAL TENDENCY
(Averages of First Order)

INTRODUCTION:

- Central tendency is defined as the tendency of the data to concentrate towards the central or middle most region of the distribution.
- In other words, Central Tendency indicates average.
- Any average is a representative value of the entire distribution value
- Average discovers uniformity in variability.
- The tendency of the variables to accumulate at the center of the distribution (data) is known as measures of central tendency.
- Measures are popularly also known as averages.



The criteria for Ideal Measures of Central Tendency

1. It should be simple to understand. (Mean, Median & Mode are easy to compute)
2. It should be based on all the observations. (AM,GM,HM are based on all the observations)
3. It should be rigidly defined (except Mode).

4. It should not be affected by extreme values (Median & Mode are not affected by extreme values.

5. It should have sampling stability or it should not be affected by sampling fluctuations. (A.M, G.M, H.M. not affected).

6. It should be capable of further algebraic treatment. (AM,GM,HM)

ARITHMETIC MEAN

- It is the best measure of central tendency and most commonly used measure
- The only drawback of this measure is that it gets highly affected by presence of extreme values in the distribution.
- Calculation of AM

1. For Simple series: A.M. $= \bar{x} = \frac{\sum x}{n}$

2. For simple frequency distribution :

Let $x_1, x_2, x_3, \ldots\ldots, x_n$ be a series, occuring with frequency $f_1, f_2, f_3, \ldots\ldots, f_n$ respectively, then

$$\text{A.M.} = \bar{x} = \frac{\sum fx}{N} = \frac{\sum fx}{\sum f} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots\ldots + f_n}; \text{ N = Total Frequency}$$

3. For Grouped Frequency Distribution:

a) Direct Method

$$\text{A.M.} = \bar{x} = \frac{\sum fx}{N} = \frac{\sum fx}{\sum f} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots\ldots + f_n}$$

Where, x = mid - values or class marks

b) Method of Assumed Mean using Step Deviation (By changing of origin and scale)

$$AM = \bar{x} = A + \left(\frac{\sum fd}{\sum f}\right) \cdot i \qquad \bullet\ d = \frac{x - a}{i}$$

Where,

X = mid-values or original values if it is a discreet series

a = Assumed Mean i.e., a value arbitrarily chosen from mid-values or any other values

I = class width or any arbitrary value

PROPERTIES

1. If all values of the variable are constant, then AM is constant.

2. $\bar{x} = \frac{\sum x}{n}$; Thus, Sum of the observations = (no. of observations) x (average).

3. Sum of deviations of values from their arithmetic mean is always zero.

4. When the values of x are equi-distant, then AM = $\frac{\text{First value} + \text{Last value}}{2}$

5. If the frequencies of variable increases or decreases by the same proportion, the value of AM will remain unaltered.

6. Weighted AM of first "n" natural numbers, when the values are equal to their corresponding weights, will be given by $\bar{x} = \frac{2n+1}{3}$

7. Sum of squares of deviation is minimum when the deviation is taken from AM.

8. AM is dependent on the change of origin and scale.

   If Y = a ± bx,

   then, $\bar{Y} = a \pm b\bar{x}$

9. Formula for calculating Combined Mean is given by: $\overline{x_c} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2}}{n_1 + n_2}$

   Where,

   $\bar{x}_1$ = mean of the first group

   $\bar{x}_2$ = mean of the second group

   n_1 = number of samples in the first group

   n_2 = number of samples in the second group

GEOMETRIC MEAN (GM)

1. *Let* $x_1, x_2, x_3, \ldots\ldots, x_n$ be a simple series, then G.M. $= \sqrt[n]{x_1 . x_2 . x_3 \ldots\ldots x_n}$ (n^{th} root of the product)

2. *Let* $x_1, x_2, x_3, \ldots\ldots, x_n$ be a series, occuring with frequency $f_1, f_2, f_3, \ldots\ldots, f_n$ respectively, then

   G.M. $= \sqrt[N]{x_1^{f_1} . x_2^{f_2} . x_3^{f_3} \ldots\ldots x_n^{f_n}}$

3. $(G.M)^n$ = Product of the observation

4. It is capable of further algebraic treatment.

5. It is less affected by sampling fluctuations compare to mode and median.

6. It is less affected by extreme values compare to AM.

7. GM cannot be calculated if any variable assumes value 0 or negative value.

8. GM is particularly useful in cases where we have to find out average rates or ratios of quantities which are changing at a cumulative rate, i.e., the change is related to the immediate preceding data. For example, average rate of depreciation by WDV method or average rate of growth of population.

9. GM is extensively used in the construction of index numbers.

10. GM is the most difficult average to calculate and understand because it involves the knowledge of logarithms.

11. Logarithm of GM of "n" observations is equal to the AM of the logarithm of these "n" observations.

12. GM is based on all observations

13. If all the observations assumed by a variable constant, say K, then the GM of the observations is also K

14. GM of the product of two variables is the product of their GM's i.e.,
if z = xy,
then GM of z = (GM of x) . (GM of y)

15. GM of the ratio of two variables is the ratio of GM's of two variables i.e.,
if z = x/y
then GM of $z = \frac{\text{GM of x}}{\text{GM of y}}$

16. Combined GM: $G_{12} = [G_1^{n_1} \; G_2^{n_2}]^{\frac{1}{n_1+n_2}} \quad \therefore \log G_{12} = \frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2}$

HARMONIC MEAN (HM)

1. Let $x_1, x_2, x_3, ..., x_n$ be a simple series, then H.M. $= \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + ... + \frac{1}{x_n}}$

2. Let $x_1, x_2, x_3, ..., x_n$ be a series, occuring with frequency $f_1, f_2, f_3, ..., f_n$ respectively, then

$$\text{H.M.} = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \frac{f_3}{x_3} + ... + \frac{f_n}{x_n}}$$

3. HM cannot be calculated if any variable assumes value 0, as inverse of 0 is undefined.

4. HM has a very restricted use, and they are usually used for calculating average speed, average rates of quantities, etc.

5. It is based on all the values.

6. It is capable of further algebraic treatment.

7. It is less affected by extreme values and sampling fluctuations compare to AM and GM.

8. If y = ax then
HM(y) = a HM (x) | GM(y) = a GM (x)

9. If all the observations are constant, HM is constant

10. Combined H.M: $H_{12} = \frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$

RULE FOR USING AM AND HM

When the average to be calculated is of the form a/b, where a and b are different quantities then

i. Use HM when 'a' is constant

ii. Use AM when 'b' is constant

For eg,

Avg. speed = ?Distance = same (given)

Use H. M

Avg. speed = ?Time = same (given)

Use A. M

we know that Speed $= \frac{\text{Distance}}{\text{Time}}$

RELATION BETWEEN AM, GM & HM

1. If the values are equal,

   AM = GM = HM.

2. If the values are distinct,

   AM > GM > HM.

3. $G^2 = A.M \times H.M.$

   $G = \sqrt{A.M. \times H.M.}$

MEDIAN:

1. Median is defined as the positional average and is regarded as the second best average after arithmetic mean.

2. Median is suitable when there is a wide range of variation in data or distribution pattern is to be studied at a varying level.

3. Median is suitable for qualitative data.

4. Median is suitable for distributions with open ends.

5. Median can be located graphically using Cumulative Frequency Polygon or Ogives.

6. The absolute sum of deviations is minimum when the deviations are taken from Median, and this property of Median is known as "Minimal Property".

7. Median is dependent on change of Origin & Scale.
If $Y = a \pm bx$
Then, $Me(Y) = a \pm bMe(x)$

Calculation

For Simple Series

Median = value corresponding to (n + 1)/2th term in the distribution

Note 1: Arrange the data in the ascending or descending order

Note 2: If the value of (n+1)/2th term is a fraction then the average of the values within which it is lying is the median.

Note 3: If n is odd median = simply the middle most value and if n is even median = average of 2 mid values

For Simple Frequency Distribution:

Median = value corresponding to the (N+1)/2th Term in the 'less than' type Cumulative Frequency column where,

N = Total Frequency

For Grouped Frequency Distribution:

$$\text{Median} = l_1 + \left[\frac{\frac{N}{2} - F}{f_m}\right] i$$

l_1 = Lower boundary of the median class i.e., the class where Cumulative Frequency N/2 falls

N = Total frequency

F = Cumulative frequency of the pre-median class.

f_m = Frequency of the median class

i = Width of the median class

MODE

1. Mode is that value of the distribution which occurs with highest frequency.

2. Mode is a crude method of finding out average and it provides only a Bird's Eye view of the distribution.

3. It is the most unstable average and the quickest method of finding out the average where we need to find out the most common value of the distribution

4. It is not affected by extreme values but it is more affected by sampling fluctuations compare to AM, GM, HM.

5. In case when distribution is Multimodal, mode is ill-defined

6. Mode is dependent on the change of origin and scale

7. If $y = a \pm bx$ then, $Mo(y) = a \pm b\, Mo(x)$

8. Mode can be located graphically using Histogram or Area Diagram or Frequency Diagram.

9. Mode does not take into account all of the observations.

10. When the classes are of unequal width, we consider frequency densities instead of class frequency to locate mode,
where frequency density = $\frac{\text{Class Frequency}}{\text{Width of the Class}}$

Calculation of Mode for Simple Series:

1. For simple series, there is no mode as all values occur with frequency = 1, i.e., same frequency.

2. For simple frequency distribution Mode can be calculated by mere inspection. The variable occurring with the highest frequency is the mode of the distribution. A distribution can be uni-modal or bi-modal, but not multi-modal.

- If only one value of variable occurs with the highest frequency, then there is only one mode.
- If two values of variable occurs with the same highest frequency, then there are two modes.
- If all values of variable occurs with same frequency, then there is no mode.
- If more than two values of variable occurs with same highest frequency, then also there is no mode.

Calculation of Mode for Grouped Frequency Distribution:

$$\text{Mode} = l_1 + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2} \right) i$$

L_1	=	Lower boundary of the modal class i.e., the class with highest frequency.
f_m	=	Frequency of the modal class
f_1	=	Frequency of the pre-modal class
f_2	=	Frequency of the post-modal class
i	=	Class width

CONCEPT OF SYMMETRICAL & ASYMMETRICAL DISTRIBUTION:

1. When in a distribution all the measures of central tendencies are equal, the distribution is said to be symmetrical.

2. For symmetrical distribution; Mean = Median = Mode.

3. Any deviation from this symmetry makes the distribution asymmetrical or skewed.

4. For moderately skewed distribution: Mean – Mode = 3(Mean – Median)

OTHER PARTITION VALUES (FRACTILES)

Partition values divides distribution in equal parts.

- QUARTILES
  - There are 3 quartiles (Q_1, Q_2, Q_3), which divides the distribution in 4 equal parts representing 25%, 50% and 75% of the data respectively.

- Q_2 is nothing but the median of the data.

- For symmetrical data, Q_2 is simple average of the extreme quartiles Q_1 (lower quartile) and Q_3 (upper quartile).

- **DECILES**
  - There are 9 deciles (D_1, D_2, ..., D_9), which divides the distribution in 10 equal parts representing 10%, 20% ..... 90% of the data respectively.

  - D_5 is nothing but the median of the data.

- **PERCENTILES**
  - There are 99 percentiles (P_1, P_2, ..., P_{99}), which divides the distribution in 100 equal parts representing 1%, 2% ..... 99% of the data respectively.

  - P_{50} is nothing but the median of the data

- **NOTE**
  - All partition values are dependent on the change of Origin and Scale.

  - All partition values can be calculated graphically through Cumulative Frequency Polygon or ogives.

Calculation of Partition Values

Type of Series	Quartiles	Deciles	Percentiles
Simple Series	$Q_i = i\left(\frac{n+1}{4}\right)$ $i = 1,2,3$	$D_i = i\left(\frac{n+1}{10}\right)$ $i = 1,2,3,.....,9$	$P_i = i\left(\frac{n+1}{100}\right)$ $i = 1,2,3,.....,99$
Simple Frequency Dist	Q_i = *value* correspo--nding to CF; i$\left(\frac{N+1}{4}\right)$	D_i = *value* correspo--nding to CF; i$\left(\frac{N+1}{10}\right)$	P_i = *value* correspo--nding to CF; i$\left(\frac{N+1}{100}\right)$
Group Frequency Dist	$Q_i = l_1 + \left[\frac{\frac{iN}{4} - f}{f_q}\right] i$	$D_i = l_1 + \left[\frac{\frac{iN}{10} - f}{f_d}\right] i$	$P_i = l_1 + \left[\frac{\frac{iN}{100} - f}{f_p}\right] i$

CLASSWORK SECTION

AIRHTEMATIC MEAN

1. Find the arithmetic mean of the numbers 3, 5, 7, 9, ... 47.
   a) 24 b) 25 c) 27 d) 31

2. Mean of 0, 3, 5, 6, 7, 9, 12, 0, 2 is:
   a) 4.89 b) 5.7 c) 5.6 d) 6.5

3. Find the arithmetic mean of the natural numbers from 1, 2, 3, ..., n.
   a) n b) (n + 1) c) (n + 1)/2 d) None of the above

4. If the AM of 3, 5, x, 12, 17 be 9, find the value of x.
   a) 5 b) 6 c) 7 d) 8

5. The arithmetic mean of 8, 1, 6 with weights 3, 2, 5 respectively is:
   a) 5 b) 5.6 c) 6 d) 4.6

6. If a variable assumes the values 1, 2, 3, 4, 5 with frequencies 1, 2, 3, 4, 5 respectively, find its arithmetic mean.
   a) 4.5 b) 4 c) 5 d) 3.67

7. If there are two groups containing 30 and 20 observation and having 50 and 60 as arithmetic means, then the combined arithmetic mean is:
   a) 52 b) 54 c) 55 d) 56

8. The average weight of students in a class of 35 students is 40 kg. If the weight of the teacher be included, the average rises by (1/2) kg; the weight of the teacher is :
   (a) 40.5 kg (b) 50 kg (c) 41 kg (d) 58 kg

9. The average salary of a group of unskilled workers is ₹ 10,000 and that of a group of skilled workers is ₹ 15,000. If the combined salary is ₹ 12,000, then what is the percentage of skilled workers?
   (a) 40% (b) 50% (c) 60% (d) none of these

10. If the relationship between two variables u and v are given by 2u + v + 7 = 0, and if AM of u is 10, then the AM of v is:

a) 17 b) −17 c) 27 d) − 27

GEOMETRIC MEAN

11. Find the Geometric mean of : 8, 4, 2.

a) 2 b) 4 c) 8 d) None of the above

12. Find the GM of the following: 3, -2, 4, 0, 5.

a) 2 b) 3 c) 1 d) cannot be determined

13. If GM of x is 10 and GM of y is 15, then the GM of xy is:

a) 150 b) Log 10 x Log 15

c) Log 150 d) None of these

14. The interest paid on the same sum yielding 3%, 4%, and 5% compound interest for 3 consecutive year respectively. What is the average yield percent on the total sum invested.

(a) 3.83% b) 4.83% c) 2.83% d) 3.99%

15. What is the GM for the numbers 8, 24 and 40?

(a) 24 (b) 12 (c) $8 \times \sqrt[3]{15}$ (d) 10

HARMONIC MEAN

16. The harmonic mean for the numbers 2, 3, 5 is

(a) 2.00 (b) 3.33 (c) 2.90 (d) $-\sqrt[3]{30}$

17. What is the HM of 1,1/2, 1/3,... ....1/n?

(a) n (b) 2n (c) $\frac{2}{(n+1)}$ (d) $\frac{n(n+1)}{2}$

18. An aeroplane flies from A to B at the rate of 500 km/hour and comes back from B to A at the rate of 700 km/hour. The average speed of the aeroplane is

(a) 600 km. per hour (b) 583.33 km. per hour

(c) 100 $\sqrt{35}$km. per hour (d) 620 km. per hour.

19. If there are two groups with 75 and 65 as harmonic means and containing 15 and 13 observation then the combined HM is given by

(a) 65 (b) 70.36 (c) 70 (d) 71.

COMBINED PROPERTIES OF AM, GM AND HM

20. If the AM and HM for two numbers are 5 and 3.2 respectively then the GM will be

(a) 16.00 (b) 4.10 (c) 4.05 (d) 4.00.

21. If the AM and GM for 10 observations are both 15, then the value of HM is

(a) Less than 15 (b) More than 15

(c) 15 (d) Can not be determined

22. If the AM and GM for two numbers are 6.50 and 6 respectively then the two numbers

(a) 6 and 7 (b) 9 and 4 (c) 10 and 3 (d) 8 and 5.

MEDIAN

23. Find the median of the following numbers: 2, 5, 8, 4, 9, 6, 7.

a) 9 b) 6 c) 8 d) None of the above

24. Find the median of the following numbers: 5, 8, 6, 9, 11, 4.

a) 6 b) 7 c) 8 d) None of the above

25. Calculate median for the following data :

No. of students	6	4	16	7	8	2
Marks	20	9	25	50	40	80

a) 20 b) 25 c) 35 d) 28

26. Two variables x and y are given by y = 2x − 3. If the median of x is 20, what is the median of y?

a) 20 b) 37 c) 40 d) 35

PARTITION VALUE

27. What is the value of the first quartile for observations 15, 18, 10, 20, 23, 28, 12, 16?
(a) 17 (b) 16 (c) 12.75 (d) 12

28. The third decile for the numbers 15, 10, 20, 25, 18, 11, 9, 12 is
(a) 13 (b) 10.70 (c) 11 (d) 11.50

MODE

29. The mode of the following observations is: 4, 3, 2, 5, 3, 4, 5, 3, 7, 3, 2, 6
a) 2 b) 3 c) 4 d) 6

30. If x and y are related by x−y−10 = 0 and mode of x is known to be 23, then the mode of y is
(a) 20 (b) 13 (c) 3 (d) 23.

COMBINED PROPERTIES OF AM, MEDIAN AND MODE

31. If arithmetic mean is 26.8, median is 27.9, then what is the value of mode?
a) 29 b) 30.1 c) 31.1 d) 29.9

32. If the Mean and Mode of a certain set of numbers be 60.4 and 50.2 respectively, find approximately the value of the Median.
a) 55 b) 56 c) 57 d) 58

MISCELLANEOUS SUM

33. What is the value of mean and median for the following data:

Marks:	5-14	15-24	25-34	35-44	45-54	55-64
No. of Students:	10	18	32	26	14	10

(a) 30 and 28 (b) 29 and 30
(c) 33.68 and 32.94 (d) 34.21 and 33.18

34. The mean and mode for the following frequency distribution

Class interval :	350-369	370-389	390-409	410-429	430-449	450-469
Frequency:	15	27	31	19	13	6

are

(a) 400 and 390 (b) 400.58 and 390

(c) 400.58 and 394.50 (d) 400 and 394.

35. The third quartile and 65th percentile for the following data are

Profits in '000 :	less than 10	10–19	20–29	30–39	40–49	50–59
No. of firms:	5	18	38	20	9	2

(a) ₹ 33,500 and ₹ 29,184 (b) ₹ 33,000 and ₹ 28,680

(c) ₹ 33,600 and ₹ 29,000 (d) ₹ 33,250 and ₹ 29,250.

36. Following is an incomplete distribution having modal mark as 44

Marks :	0–20	20–40	40–60	60–80	80–100
No. of students :	5	18	?	12	5

What would be the mean marks?

(i) 45 (ii) 46 (iii) 47 (iv) 48

37. For the following incomplete distribution of marks of 100 pupils, median mark is known to be 32.

Marks:	0–10	10–20	20–30	30–40	40–50	50–60
No. of Students:	10	-	25	30	-	10

What is the mean mark?

(a) 32 (b) 31 (c) 31.30 (d) 31.50

THEORETICAL ASPECTS

38. Measures of central tendency for a given set of observations measures

(a) The scatterness of the observations

(b) The central location of the observations

(c) Both (a) and (b)

(d) None of these.

39. While computing the AM from a grouped frequency distribution, we assume that
    (a) The classes are of equal length
    (b) The classes have equal frequency
    (c) All the values of a class are equal to the mid-value of that class
    (d) None of these.

40. Which of the following statements is wrong?
    (a) Mean is rigidly defined
    (b) Mean is not affected due to extreme values.
    (c) Mean has some mathematical properties
    (d) All these

41. Which of the following statements is true?
    (a) Usually mean is the best measure of central tendency
    (b) Usually median is the best measure of central tendency
    (c) Usually mode is the best measure of central tendency
    (d) Normally, GM is the best measure of central tendency

42. For open-end classification, which of the following is the best measure of central tendency?
    (a) AM (b) GM (c) Median (d) Mode

43. The presence of extreme observations does not affect
    (a) AM (b) Median (c) Mode (d) (b) and (c) both

44. In case of an even number of observations which of the following is median?
    (a) Any of the two middle-most value
    (b) The simple average of these two middle values
    (c) The weighted average of these two middle values
    (d) Any of these

45. The most commonly used measure of central tendency is
    (a) AM (b) Median (c) Mode (d) Both GM and HM.

46. Which one of the following is not uniquely defined?
    (a) Mean (b) Median (c) Mode (d) All of these measures

47. Which of the following measure of the central tendency is difficult to compute?

(a) Mean (b) Median (c) Mode (d) GM

48. Which measure(s) of central tendency is(are) considered for finding the average rates?

(a) AM (b) GM (c) HM (d) Both (b) and (c)

49. For a moderately skewed distribution, which of he following relationship holds?

(a) Mean − Mode = 3 (Mean − Median)

(b) Median − Mode = 3 (Mean − Median)

(c) Mean − Median = 3 (Mean − Mode)

(d) Mean − Median = 3 (Median − Mode)

50. Weighted averages are considered when

(a) The data are not classified

(b) The data are put in the form of grouped frequency distribution

(c) All the observations are not of equal importance

(d) Both (a) and (c).

51. Which of the following results hold for a set of distinct positive observations?

(a) AM ≥ GM ≥ HM (b) HM ≥ GM ≥ AM

(c) AM > GM > HM (d) GM > AM > HM

52. When a firm registers both profits and losses, which of the following measure of central tendency cannot be considered?

(a) AM (b) GM (c) Median (d) Mode

53. Quartiles are the values dividing a given set of observations into

(a) Two equal parts (b) Four equal parts

(c) Five equal parts (d) None of these

54. Quartiles can be determined graphically using

(a) Histogram (b) Frequency Polygon

(c) Ogive (d) Pie chart.

55. Which of the following measure(s) possesses (possess) mathematical properties?

(a) AM (b) GM (c) HM (d) All of these

56. Which of the following measure(s) satisfies (satisfy) a linear relationship between two variables?
(a) Mean (b) Median (c) Mode (d) All of these

57. Which of the following measures of central tendency is based on only fifty percent of the central values?
(a) Mean (b) Median (c) Mode (d) Both (a) and (b)

58. In the formula, Mode = $L_1 + (d_1 \times c)/(d_1 + d_2)$, d_1 is the difference of frequencies in the modal class & the _____ class.
(a) preceding (b) following (c) both (d) none

59. In the formula, Mode = $L_1 + (d_1 \times c)/(d_1 + d_2)$, d_2 is the difference of frequencies in the modal class & the ———— class.
(a) preceding (b) succeeding (c) both (d) none

60. ——— of a set of observations is defined to be their sum, divided by the no. of observations.
(a) H.M (b) G.M (c) A.M (d) none

61. The sum of the squares of deviations of a set of observations has the smallest value, when the deviations are taken from their
(a) A.M (b) H.M (c) G.M (d) none

62. ——— is equal to the value corresponding to cumulative frequency $3(N + 1)/4$ from simple frequency distribution
(a) Median (b) 1^{st} quartile (c) 3^{rd} quartile (d) 1^{st} decile

63. ——— is equal to the value corresponding to cumulative frequency $k(N + 1)/10$ from simple frequency distribution
(a) Median (b) k^{th} decile
(c) k^{th} percentile (d) none

64. For 899, 999, 391, 384, 590, 480, 485, 760, 111, 240
Rank of median is
(a) 2.75 (b) 5.5 (c) 8.25 (d) none

65. The deviations from median are ——————— if negative signs are ignored as compared to other measures of central tendency.

(a) minimum (b) maximum (c) same (d) none

66. The average discovers

(a) uniformity in variability (b) variability in uniformity of distribution

(c) both (d) none

ANSWERS - SUMS				ANSWERS - THEORITICAL ASPECTS			
Q. No.	Ans	Q. No.	Ans	Q. No.	Ans	Q. No.	Ans
1	b	21	c	41	a	61	a
2	a	22	b	42	c	62	c
3	c	23	b	43	d	63	b
4	d	24	b	44	b	64	b
5	b	25	b	45	a	65	a
6	d	26	b	46	c	66	a
7	b	27	c	47	d		
8	d	28	b	48	d		
9	a	29	b	49	a		
10	d	30	b	50	c		
11	b	31	b	51	c		
12	d	32	c	52	b		
13	a	33	c	53	b		
14	d	34	c	54	c		
15	c	35	a	55	d		
16	c	36	d	56	d		
17	c	37	c	57	b		
18	b	38	b	58	a		
19	c	39	c	59	b		
20	d	40	b	60	c		

HOMEWORK SECTION

1. If x and y are related by x − y − 10 = 0 and mode of x is known to be 23, then the mode of y is

   (a) 20 (b) 13 (c) 3 (d) 23

2. A man travels at a speed of 20 km/hr and then returns at a speed of 30 km/hr. His average speed of the whole journey is:

   (a) 25 km/hr (b) 24.5 km/hr

   (c) 24 km/hr (d) none

3. The median of the data 13, 8, 11, 6, 4, 15, 2, 18, is

   (a) 5 (b) 8 (c) 11 (d) 9.5

4. The sum of the squares of deviations of a set of observations has the smallest value, when the deviations are taken from their

   (a) A.M. (b) H.M. (c) G.M. (d) None

5. Which of the following result hold for a set of distinct positive observations?

   (a) A.M. ≥ G.M. ≥ H.M.

   (b) G.M. > A.M. > H.M.

   (c) G.M. ≥ A.M. ≥ H.M.

   (d) A.M. > G.M. > H.M.

6. If the A.M. and H.M. for two numbers are 5 and 3.2 respectively then the G.M. will be

   (a) 4.05 (b) 16 (c) 4 (d) 4.10

7. An aeroplane flies from A to B at the rate of 500 km/hr and comes back from B to A at the rate of 700 km/hr. The average speed of the aeroplane is:

   (a) 600 km/hr (b) 583.33 km/hr

   (c) $100\sqrt{35}$ km/hr (d) 620 km/hr

8. For a moderately skewed distribution, which of the following relationship holds?
   (a) Mean − Median = 3 (Mean − Mode)
   (b) Median − Mode = 3 (Mean − Median)
   (c) Mean − Mode = 3 (Mean − Median)
   (d) Mean − Median = 3 (Mean − Mode)

9. _____ & _____ are called ratio averages:
   (a) H.M. & G.M. (b) H.M. & A.M.
   (c) A.M. & G.M. (d) None

10. Extreme values have _____ effect on mode.
   (a) High (b) Low
   (c) No (d) None of these

11. The mean salary for a group of 40 female workers is ₹ 5200 per month and that for a group of 60 male workers is ₹ 6800 per month. What is the combined salary?
   (a) ₹ 6160 (b) ₹ 6280
   (c) ₹ 6890 (d) ₹ 6920

12. If there are two groups with 75 and 65 as harmonic means and containing 15 and 13 observations, then the combined H.M. is given by:
   (a) 70 (b) 80 (c) 70.35 (d) 69.48

13. The G.M. of 4, 6 and 8 is:
   (a) 4.77 (b) 5.32 (c) 6.14 (d) 5.77

14. G.M. is a better measure than others when
   (a) ratios and percentages are given
   (b) interval of scale is given
   (c) Both (a) and (b)
   (d) Either (a) or (b)

15. The median of $x, \frac{x}{2}, \frac{x}{3}, \frac{x}{5}$ is 10. Find where $x > 0$
   (a) 24 (b) 32 (c) 8 (d) 16

16. The average salary of 50 men was ₹ 80 but it was found that salary of 2 of them were ₹ 46 and ₹ 28 which was wrongly taken as ₹ 64 and ₹ 82. The revised average salary is:

(a) ₹ 80 (b) ₹ 78.56 (c) ₹ 85.26 (d) ₹ 82.92

17. If A be the A.M. of two positive unequal quantities X and Y and G be their G.M. then:

(a) $A < G$ (b) $A > G$

(c) $A \leq G$ (d) $A \geq G$

18. When mean is 3.57 and mode is 2.13 then the value of median is _____.

(a) 3.09 (b) 5.01

(c) 4.01 (d) None of these

19. The harmonic mean of 1, 1/2, 1/3 . . . 1/n is

(a) 1/(n + 1) (b) 2/(n + 1)

(c) (n + 1)/2 (d) 1/(n – 1)

20. The mean weight of 15 students is 110 kg. The mean weight of 5 of them is 100 kg and of another five students is 125 kg. then the mean weight of the remaining students is:

(a) 120 (b) 105

(c) 115 (d) none of these

21. In a class of 11 students, 3 students were failed in a test. 8 students who passed secured 10, 11, 20, 15, 12, 14, 26 and 24 marks respectively. What will be the median marks of the students:

(a) 12 (b) 15 (c) 13 (d) 13.5

22. A lady travel at a speed of 20 km/h and returned at quicker speed. If her average speed of the whole journey is 24 km/hr, find the speed of return journey (in km/h)

(a) 25 (b) 30 (c) 35 (d) 38

23. Let the mean of the variable 'x' be 50, then the mean of u = 10 + 5x will be :

(a) 250 (b) 260 (c) 265 (d) 273

24. If the difference between mean and Mode is 63, then the difference between Mean and Medium will be _____

(a) 63 (b) 31.5 (c) 21 (d) None of the above

25. If the Arithmetic mean between two numbers is 64 and the Geometric mean between them is 16. The Harmonic Mean between them is _____

(a) 64 (b) 4 (c) 16 (d) 40

26. The average of 5 quantities is 6 and the average of 3 is 8 what is the average of the remaining two.

(a) 4 (b) 5 (c) 3 (d) 3.5

27. The median of following numbers, which are given is ascending order is 25. Find the value of X.

11, 13, 15, 19, (x + 2), (x + 4), 30, 35, 39, 46

(a) 21 (b) 20 (c) 15 (d) 30

28. The average age of a group of 10 students was 20 years. The average age increased two years when two new students joined the group. What is the average age of two new students who joined the group?

(a) 22 years (b) 30 years
(c) 20 years (d) 32 years

29. Geometric Mean of three observations 40, 50 and X is 10. The value of X is

(a) 2 (b) 4
(c) 1/2 (d) None of the above

30. The mean of first three term is 14 and mean of next two terms is 18. Then mean of all five term is:

(a) 14.5 (b) 15 (c) 14 (d) 15.6

31. The mean salary of a group of 50 persons is ₹ 5850. Later on it is discovered that the salary of one employee has been wrongly taken as ₹ 8000 instead of ₹ 7800. The corrected mean salary is

(a) ₹ 5854 (b) ₹ 5846
(c) ₹ 5650 (d) None of the above

32. If the mode of a data is 18 and mean is 24, then median is

(a) 18 (b) 24 (c) 22 (d) 21

33. The point of intersection of the "less than" and "more than" ogives correspond to

(a) Mean (b) Mode
(c) Median (d) 10th Percentile

34. A man travels from Agra to Gwalior at an average speed of 30 km per hour and back at an average speed of 60 km per hour. What is his average speed?

(a) 38 km per hour (b) 40 km per hour
(c) 45 km per hour (d) 35 km per hour

35. Which of the following measures of central tendency cannot be calculated by graphical method?

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

36. Geometric mean of 8, 4, 2 is

(a) 4 (b) 2 (c) 8 (d) none of these

37. The average age of 15 students of a class is 15 years. Out of them, the average age of 5 students is 14 years and that of the other 9 students is 16 years. The age of the 15th student is:

(a) 11 years (b) 14 years
(c) 15 years (d) None of these

38. Averages whose value can be determined graphically?

(a) Mode, Median (b) Mean, Mode
(c) Mean, Median (d) None of the above

39. Which of the following statements is true?

(a) Median is based on all the observations
(b) The mode is the mid value
(c) The median is the second quartile
(d) The mode is the fifth decile

40. The mean of the following data is 6. Find the value of 'P'

x	2	4	6	10	P + 5
f	3	2	3	1	2

(a) 4 (b) 6 (c) 8 (d) 7

41. The third decile for the numbers 15, 10, 20, 25, 18, 11, 9, 12, is :

(a) 13 (b) 10.70 (c) 11 (d) 11.50

42. A random variable X has uniform distribution on the interval (−3, 7). The mean of the distribution is:

(a) 2 (b) 4 (c) 5 (d) 6

43. If the arithmetic mean of two numbers is 10 and the geometric mean of these numbers is 8, then the harmonic mean is:

(a) 9 (b) 8.9 (c) 6.4 (d) None of these

44. The harmonic mean H of two numbers is 4 and their arithmetic mean A and the geometric mean G satisfy the equation $2A + G^2 = 27$, then the numbers are

(a) (1, 3) (b) (9, 5) (c) (6, 3) (d) (12, 7)

45. Quartiles can be determined graphically using:

(a) Histogram (b) Frequency polygon
(c) Ogive curve (d) Pie chart

46. In a class of 50 students, 10 have failed and their average marks in 2.5. The total marks secured by the entire class were 281. The average marks who have passed is:

(a) 5.32 (b) 7.25
(c) 6.40 (d) None of the above

47. If the mean of two numbers is 30 and geometric mean is 24 then what will be these two numbers?

(a) 36 and 24 (b) 30 and 30
(c) 48 and 12 (d) None of these

48. For moderately skewed distribution of marks in commerce for a group of 200 students the mean marks and mode marks were found to be 55.60 and 46. What is the median marks?

(a) 55.5 (b) 60.5 (c) 52.4 (d) None of these

49. Mean for the data 6, 4, 1, 6, 5, 10, 3 is 5 when each observation added by 2, what is mean of the data

(a) 5 (b) 6 (c) 7 (d) 10

50. The average of 10 observations is 14.4. If the average of first 4 observations is 16.5. The average of remaining 6 observations is:

(a) 13.6 (b) 13.0 (c) 13.2 (d) 12.5

51. The ordering of a particular design of a cloth show room, a _____ size be more appropriate

(a) median (b) mean (c) mode (d) all of these

52. The rates of returns from three different shares are 100%, 200% and 400% respectively. The average rate of return will be :

(a) 350% (b) 233.33% (c) 200% (d) 300%

53. If geometric mean is 6 and arithmetic mean is 6.5, then harmonic mean will be:

(a) $\frac{6^2}{6.5}$ (b) $\frac{6}{6.5^2}$

(c) $\frac{6}{6.5}$ (d) None of the above

54. A company's past 10 years average earning is ₹ 40 crores. To have the same average earning for 11 years including these 10 years, how much earning must be made by the company in the eleventh year?

(a) ₹ 40 crores (b) ₹ $\frac{40 \times 10}{11}$ crores

(c) More than ₹ 40 crores (d) None of these

55. A person purchases 5 rupees worth of eggs from 100 different markets. You are to find the average number of eggs per rupee purchased from all the markets taken together. The suitable average in this case is:

(a) A.M. (b) G.M. (c) H.M. (d) None of the above

56. _____ is the reciprocal of the AM of the reciprocal of observations.

(a) HM (b) GM

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

57. If the mean value of seven numbers 7, 9, 12, X, 4, 11 and 5 is 9, then the missing number X will be:

(a) 13 (b) 14 (c) 15 (d) 8

58. When all observations occur with equal frequency _____ does not exist.

(a) median (b) mode

(c) mean (d) None of the above

59. If the variables x and z are so related that $z = ax + b$ for each $x = x_1$ where a and b are constant, then $\bar{Z} = a\bar{x} + b$

(a) True (b) false

(c) both (d) none

60. If each item is reduced by 15 A.M. is

(a) reduced by 15 (b) increased by 15

(c) reduced by 10 (d) none

61. For 899, 999, 391, 384, 590, 480, 485, 760, 111, 240 Rank of median is

(a) 2.75 (b) 5.5 (c) 8.25 (d) none

62. The average of a series of overlapping averages, each of which is based on a certain number of item within a series is know as

(a) Moving average (b) Weighted average

(c) Simple average (d) None

63. The median of the data 5, 6, 7, 7, 8, 9, 10, 11, 11, 12, 15, 18 and 19 is

(a) 10.5 (b) 10 (c) 11 (d) 11.5

64. The mean of 20 items of a data is 5 and if each item is multiplied by 3, then the new mean will be

(a) 5 (b) 10 (c) 15 (d) 20

65. The Geometric mean of 3, 6, 24 and 48 is

(a) 8 (b) 12 (c) 24 (d) 6

66. The Algebraic sum of the deviation of a set of values from their arithmetic mean is

(a) > 0 (b) = 0 (c) < 0 (d) None of the above

67. Which one of the following is not a central tendency?

(a) Mean Deviation (b) Arithmetic mean

(c) Median (d) Mode

68. If total frequencies of three series are 50, 60 and 90 and their means are 12, 15 and 20 respectively, then the mean of their composite series is

(a) 16 (b) 15.5 (c) 16.5 (d) 14.5

69. If in a moderately skewed distribution the values of mode and mean are 32.1 and 35.4 respectively, then the value of the median is

(a) 34.3 (b) 33.3 (c) 34 (d) 33

ANSWER

1	b	21	a	41	b	61	b
2	c	22	b	42	a	62	a
3	d	23	b	43	c	63	b
4	a	24	c	44	c	64	c
5	d	25	b	45	c	65	b
6	c	26	c	46	c	66	b
7	b	27	a	47	c	67	a
8	c	28	d	48	c	68	c
9	a	29	c	49	c	69	a
10	c	30	d	50	b		
11	a	31	b	51	c		
12	a	32	c	52	c		
13	d	33	c	53	a		
14	a	34	b	54	a		
15	a	35	a	55	c		
16	b	36	a	56	a		
17	b	37	a	57	c		
18	a	38	a	58	b		
19	b	39	c	59	a		
20	b	40	d	60	a		

HOMEWORK SOLUTION

1. (b) $X - Y - 10 = 0$
   $Mode(X) - Mode(Y) - 10 = 0$
   $Mode(y) = Mode(X) - 10 = 23 - 10 = 13$

2. (c) Average speed is the HM of 20 and 30.
   $AS = (2*20*30)/(20+30) = 1200/50 = 24$

3. (d) Arranging the given data in ascending order: 2, 4, 6, 8, 11, 13, 15, 18
   Median is the value of $(8 + 1)/2 = 4.5^{th}$ item = ½(4^{th} item + 5^{th} item) = ½(8+11) = 9.5

4. (a) The sum of Squares of deviations is least, when it is taken from its AM.

5. (d) For set of distinct positive observations: AM > GM > HM

6. (c) $GM^2 = AM * HM = 5 * 3.2 = 16$
   $GM = 4$

7. (b) Average Speed is the HM of 500 and 700 = (2*500*700)/(500+700) = 700000/1200 = 583.33

8. (c) For a moderately skewed distribution: Mean − Mode = 3(Mean − Median)

9. (a) HM and GM are known as ratio averages.

10. (c) Extreme values have NO effect on mode.

11. (a) Combined average salary = (40*5200 + 60*6800)/(40 + 60) = 616000/100 = 6160.

12. (a) Combined AM = (15*1/75 + 13*1/65)/(15+13) = 0.4/28
    Combined HM = 28/0.4 = 70.

13. (d) (Find the cube of the options, the option which gives 192 is the answer)
$GM^3 = 4*6*8 = 192$
GM = 5.77

14. (a) GM is the better measure than others when ratios and percentages are given.

15. (a) Median = ½ (x/2 + x/3) = 5x/12 = 10. Thus, x = 24.

16. (b) Correct Total = (50 * 80) + 46 + 28 – 64 – 82 = 3928
Correct Mean = 3298 / 50 = 78.56

17. (b) For two unequal quantities: A > G

18. (a) Mean – Mode = 3 (Mean – Median)
3Median = 2 Mean + Mode = 2*3.57 + 2.13 = 9.27
Median = 9.27/3 = 3.09

19. (b) HM = n / (1/1 + 1/1/2 + 1/1/3 + ... + 1/1/n) = n/(1 + 2 + 3 + ... + n) = 2/(n+1)

20. (b) Mean of remaining 5 = 1/5 [15*110 – 5*100 – 5*125] = 1/5*525 = 105.

21 (a) Arranging the marks in ascending order: A B C 10 11 12 14 15 20 24 26
Median is the marks of (11 + 1)/2 = 6th student from any corner = 12

22. (b) 24 = 2*20*S / (20 + S)
480 + 24S = 40S
16S = 480
S = 480/16 = 30

23. (b) Mean (U) = 10 + 5.Mean(X) = 10 + 5*50 = 260

24. (c) Mean – Mode = 63 = 3(Mean – Median)
(Mean – Median) = 63/3 = 21

25. (b) $GM^2 = AM * HM$

$16^2 = 64 * HM$

$HM = 256/64 = 4$

26. (c) Average of remaining 2 = ½ [5*6 – 3*8] = ½ [30 – 24] = 6/2 = 3

27. (a) 11, 13, 15, 19, (x + 4), 30, 35, 39, 46

Median is the 5^{th} value

Rank = $\frac{N+1}{2}$ $\frac{10}{5}$ 5^{th}

x+ 4= 25

x = 25 – 4

= 21

28. (d) Average age of two new person = ½ [12*22 – 10*20] = ½ (264 – 200) = 32

29. (c) 103 = 40 * 50 * X

X = 1000/2000 = ½

30. (d) Mean = 1/5 [3*14 + 2*18] = 1/5(42 + 36) = 78/5 = 15.6

31. (b) Correct Mean = 1/50 [50*5850 – 8000 + 7800] = 1/50 [292300] = 5846

32. (c) Mean – Mode = 3(Mean – Median)

3Median = 2Mean + Mode = 2*24 + 18 = 66

Median = 66/6 = 22.

33. (c) The point of intersection of the "less than" and "more than" Ogives correspond to the Median of the distribution.

34. (b) Average Speed = (2*30*60)/(30+60) = 3600/90 = 40

35. (a) Mean cannot be calculated graphically.

36. (a) $GM^3 = 8 * 4 * 2 = 64 = 4^3$

GM = 4.

37. (a) The age of 15^{th} student = 15*15 – 5*14 – 9*16 = 225 – 70 – 144 = 11

38. (a) Graphically Median and Mode of a distribution can be calculated.

39. (c) Median is not based on all observations. Mode is the most popular value and not the mid-value. Median is the second quartile.

40. (d) $\sum fx = 2*3 + 4*2 + 6*3 + 10*1 + (p+5)*2 = 6+8+18+10+2P+10 = 2P + 52$
$\sum f = 3+2+3+1+2 = 11$
$6 = (2P + 52)/11$
$66 = 2P + 52$
$2P = 14$
$P = 7$.

41. (b) Ascending order: 9, 10, 11, 12, 15, 18, 20, 25
3^{rd} Decile = $3(8 + 1)/10 = 2.7^{th}$ element = 2^{nd} element + 0.7 ($3^{rd} - 2^{nd}$) = 10 + 0.7(11 – 10) = 10.7

42. (a) Mean = (-3 + 7)/2 = 4/2 = 2.

43. (c) $8^2 = 10 * HM$
$HM = 64/10 = 6.4$.

44. (c) $2A + G^2 = 2A + A*H = 2A + 4A = 6A = 27$. $A = 27/6 = 4.5$
$G^2 = 4 * 4.5 = 18$
First number = $4.5 + \sqrt{(4.5^2 - 18)} = 4.5 + 1.5 = 6$
Second number = $4.5 - \sqrt{(4.5^2 - 18)} = 4.5 - 1.5 = 3$

45. (c) Quartiles are calculated graphically using Ogives.

46. (c) Average marks of students who have passed = 1/40[281 – 10*2.5] = 1/40(256) = 6.4.

47. (c) $X = 30 + \sqrt{(30^2 - 24^2)} = 30 + 18 = 48$
$Y = 30 - 18 = 12$

48. (c) 3Median = 2Mean + Mode = 2*55.60 + 46 = 157.2
Median = 157.2/3 = 52.4

49. (c) New Mean = 5 + 2 = 7.

50. (b) Average of remaining 6 = 1/6 [10*14.4 – 4*16.5] = 1/6[144 – 66] = 78/6 = 13.

51. (c) Mode is an appropriate average for a cloth show room, which shall want to order that particular design, which sells more.

52. (c) Average rate = GM of rates
$= \sqrt[3]{100 \times 200 \times 400}$
$= \sqrt[3]{8 \times 1000000}$
$= \sqrt[3]{8} \times \sqrt[3]{10^6}$
$= (2^3)^{\frac{1}{3}} \quad (10^6)^{\frac{1}{3}}$
= 2 x 100
= 200 %

53. (a) $6^2 = 6.5 * H$
$H = 36/6.5 = 6^2/6.5$

54. (a) Earning in the 11th year = 40 crores. (to maintain the average of last 10 years)

55. (a) AM is to be used to find average rate at which eggs are to be purchased.

56. (a) HM is the reciprocal of the AM of the reciprocal of observations.

57. (c) (7 + 9 + 12 + X + 4 + 11 + 5) = 7*9 = 63
48 + X = 63
X = 63 – 48 = 15.

58. (b) When all observations occur with equal frequency, we can't calculate its Mode.

59. (a) If Z = aX + b, Mean(Z) = a. Mean(X) + b
(As Mean is dependent on both change of scale and change of origin)

60. (a) If each item is reduced by 15, AM also reduced by 15.

61. (b) Rank of Median = $\frac{N+1}{2} = \frac{9+1}{2} = \frac{10}{2}$ =5th
Rank of Median is 5

62. (a) Average of a series of overlapping averages, each of which is based on a certain number of item within a series is known as Moving Averages.

63. (b) 5 6 7 7 8 9 | 10 | 11 11 12 15 18 19

$m = \frac{N+1}{2} = \frac{14}{2} = 7^{th}$ value

Median = 10

64. (c) New mean = 5 * 3 = 15

65. (b) $G^4 = 3 * 6 * 24 * 48 = 20736 = 12^4$
G = 12.

66. (b) The algebraic sum of the deviation of a set of values from their AM is always zero.

67. (a) Mean Deviation is a measure of dispersion.

68. (c) Mean = (50*12 + 60*15 + 90*20) / (50+60+90) = (600+900+1800)/200 = 3300/200 = 16.5

69. (a) 3Median = 2Mean + Mode = 2*35.4 + 32.1 = 70.8 + 32.1 = 102.9
Median = 102.9/3 = 34.3

SELF ASSESSMENT TEST 2
30 Marks

1. What is the major assumption we make when computing a mean from grouped data?
   a) No value occurs more than once
   b) Each class contains exactly the same number of values
   c) All values are discrete
   d) Every value in a class is equal to the mid-point

2. Median is:
   a) 50^{th} percentile b) 2^{nd} quartile c) 5^{th} decile d) All of the above

3. Suitable average for finding out the mean size of sale of shoes shall be:
   a) Arithmetic mean b) Median c) Mode d) None of the above

4. A cyclist goes from a place to another and returns by the same route. He pedals his way uni-formly with speed U while going and with speed V while returning. The average speed of his journey is:
   a) (U + V)/2 b) 2/(U + V) c) 2UV/(U + V) d) None of the above

5. The mean of the value 1, 2, 3, ..., 120 with respective frequencies F, 2F, 3F,...., 120F is equal to:
   a) $(60 + 1)^2$ b) (240 + 1)/6 c) 60 d) 241/3

6. If the median of four numbers: 3, 4, 8 and X is 5, then what is the value of X?
   a) 11 b) 5 c) 10 d) 6

7. Find two positive numbers A and B, the AM of which is 5 and the GM is 4.
   a) 2, 8 b) 8, 2 c) 6, 4 d) (a) and (b) both

8. Which of the following statement is not CORRECT?

a) In a negatively skewed distribution, the value of mode is greater than the arithmetic mean.

b) The subtraction of a constant from each item in the data to be averaged, changes the average.

c) Geometric mean is square root of the product of all observations.

d) The division of a constant from each item in the data to be averaged, changes the average.

9. Which measure of central tendency is not affected by the extreme values?

a) Arithmetic Mean and Median

b) Median and Mode

c) Mode and Arithmetic Mean

d) Geometric Mean and Harmonic Mean

10. Of the various measure of central tendency, which of the following can be used when measurement are on an ordinal scale?

a) Arithmetic Mean and Median

b) Mode and Arithmetic Mean

c) Median and Mode

d) All of the above

11. The mean of 15 observations is 15. If the two numbers 18 and 22 are excluded, then the mean of the remaining numbers is:

a) 10.53 b) 12.49 c) 14.23 d) 15.49

12. While dividing each entry in a data by a non-zero number A, the arithmetic mean of the new data:

a) Does not change b) Increased

c) Is divided by A d) Is diminished by A

13. If arithmetic mean of two items is 16 and their harmonic mean is 9. Find their geometric mean.

a) 10 b) 12 c) 14 d) 16

14. Find the median of 33, 86, 68, 80, 48, 70, 64.

a) 80 b) 68 c) 70 d) 48

15. The mean monthly salary of all employees in a company is Rs. 25,000. The mean salaries of male and female employees are Rs. 27,000 and Rs. 17,000 respectively, the percentage of males employed by the company is:

a) 80% b) 20% c) 30% d) 70%

16. The AM of 7, (x − 2), 10, (x + 3) is 9. Find x.

a) 8 b) 9 c) 7 d) None of the above

17. Calculate the geometric mean of 3, 6, 24, 48.

a) 6 b) 12 c) 24 d) None of the above

18. Find the GM of 4, 6, 9 with weight 1, 2, 1 respectively.

a) 6 b) 12 c) 24 d) None of the above

19. If the GM of A, 4, 8 be 6, find the value of A.

a) 6.75 b) 12.25 c) 24.35 d) None of the above

20. The means of two samples of sizes 50 and 100 respectively are 54.1 and 50.3. Obtain the mean of the sample of size 150 obtained by combining the two samples.

a) 50.47 b) 51.57 c) 52.37 d) 54.48

21. Find the Harmonic Mean of the following numbers: 1. ½, 1/3, ¼.

a) 0.4 b) 0.5 c) 0.25 d) None of the above

22. Find the Mode of the following numbers: 7, 4, 3, 5, 6, 3, 3, 2, 4, 3, 4, 3, 3, 4, 4, 3, 2, 2, 4, 3, 5, 4, 3, 4, 3, 1, 2, 3.

a) 4 b) 3 c) 2 d) None of the above

23. Find the median of 88, 72, 33, 29, 70, 86, 54, 91, 61, 57.

a) 65 b) 66 c) 65.5 d) None of the above

24. In a moderately symmetric distribution the mode and mean are 12.30 and 18.48 respectively. Find the median.

a) 16 b) 16.42 c) 16.24 d) None of the above

25. The average weight of A, B, C is 45 Kg. If the average weight of A and B be 40 Kg and that of B and C be 43 Kg, then the weight of B is:

a) 15 Kg b) 20 Kg c) 25 Kg d) 31 Kg

26. The average attendance of a school for the first 3 days of the week is 325 and for the first 4 days of the week it is 320. The students present on the fourth day were:

a) 305 b) 310 c) 315 d) 325

27. The AM calculated from the following frequency distribution is known to be 67.45. Find the value of F.

Class Interval	60-62	63-65	66-68	69-71	72-74
Frequency	15	54	F	81	24

a) 120 b) 126 c) 134 d) None of the above

28. The median of the following incomplete frequency distribution is 4. Find the frequency when X = 8.

X:	1	2	3	4	5	6	7	8
CF:	2	3	4	5	6	7	8	?

a) 1 b) 2 c) 3 d) 4

29. Calculate the mode of the following frequency distribution:

Class Interval	Frequency
0 – 6	13
6 – 12	25
12 – 18	57
18 – 24	79
24 – 30	105
30 – 36	79
36 – 42	57
42 – 48	25
48 – 54	13

a) 27 b) 28 c) 29 d) 31

30. The algebraic sum of the deviations of 50 observations measured from 90 is – 110. Find the AM of the observations.

a) 90 b) 110 c) 88.7 d) 87.8

EXPLANATORY ANSWERS

1. While calculating mean from the group distribution, it is assumed that every value in a class is equal to the mid-point of that class. Option D

2. Median = 50^{th} percentile = 2^{nd} quartile = 5^{th} decile. Option D

3. Mode is the suitable average to find the mean sale of shoes. Option C

4. Average speed is the HM between U and V = 2/(1/U + 1/V) = 2UV/(U + V). Option C

5. $\sum FX = (1*F) + (2*2F) + ... + (120*120F) = F(120)(120 + 1)(240 + 1)/6$
   $\sum F = F + 2F + 3F + ... + 120F = F(120)(120 + 1)/2$
   Mean = $\sum FX / \sum F = (240 + 1)/3 = 241/3$
   Option D

6. Median of 4 numbers is 5. The sum of the central 2 values = 5*2 = 10
   If we consider 4 and 8 as the middle two numbers, its total is not 10.
   Thus 4 + X = 10. X = 6. Option D

7. $A = AM + \sqrt{(AM^2 - GM^2)} = 5 + \sqrt{(25 - 16)} = 8$
   $B = AM - \sqrt{(AM^2 - GM^2)} = 5 - \sqrt{(25 - 16)} = 2$
   Both option A and option B are correct. Thus, correct answer is Option D

8. Geometric mean is the nth root of the product of n observations given in the data set. Option C

9. Both Median and Mode are not affected by the presence of the extreme values. Option B

10. Arithmetic mean is only possible for cardinal scale. For ordinal scale all partition and positional values are possible. Option C

11. Mean of remaining = (15*15 − 18 − 22)/13 = 14.23. Option C

12. The AM of the new data is also get divided by A. Option C

13. GM^2 = AM * HM = 16 * 9 = 144. GM = 12. Option B

14. After arranging: 33, 48, 64, 68, 70, 80, 86. Median is (7 + 1)/2 = 4^{th} item from any end = 68. Option B

15. 25000(M + F) = 27000M + 17000F
    2000M = 8000F; M:F = 4:1. Percentage of males employed = 4/5 * 100 = 80%. Option A

16. 9*4 = 7 + x − 2 + 10 + x + 3 = 2x + 18; 2x = 18. x = 9. Option B

17. 3*6*24*48 = 34.28. GM = $(3^4.2^8)^{1/4}$ = $3*2^2$ = 12. Option B

18. GM = $\sqrt[4]{4^1.6^2.9^1}$ = $\sqrt[4]{1296}$ = 6. Option A

19. 6^3=Ax4x8; A = 216/32 = 6.75. Option A

20. Combined Mean = [(54.1*50) + (50.3*100)]/150 = (2705 + 5030)/150 = 51.57. Option B

21. HM = 4/(1 + 2 + 3 + 4) = 4/10 = 0.40. Option A

22. On observation, value 3 occur maximum number of times. Mode = 3. Option B

23. After arranging: 29, 33, 54, 57, 61, 70, 72, 86, 88, 91. Median is ½ (5^{th} + 6^{th} value) = ½ (61 + 70) = ½(131) = 65.5. Option C

24. 3(Mean − Median) = (Mean − Mode)
    3(18.48 − Median) = (18.48 − 12.30) = 6.18
    Median = 18.48 − 2.06 = 16.42. Option B

25. A+B+C = 45*3 = 135
A+B = 40*2 = 80
B+C = 43*2 = 86
B = (A+B) + (B+C) – (A+B+C) = 80 + 86 – 135 = 31. Option D

26. Attendance of 4th day = (320 * 4) – (325 * 3) = 1280 – 975 = 305. Option A

27. ∑FX = (15*61)+ (54*64) + (F*67) + (81*70) + (24*73) = 915+3456+67F+5670+1752
= 11793+67F
∑F = 15 + 54 + F + 81 + 24 = 174 + F
67.45 = (11793 + 67F)/(174 + F)
11736.3 + 67.45F = 11793 + 67F
0.45F = 56.7
F = 126. Option B

28. First prepare frequency column from CF column by subtracting values
Option Method
A) Missing frequency = 1
put missing frequency = 1 & apply Median Formula = (N+1)/2 th Term
Answer Matches with given data M = 4
Hence Option A

29. Mode = 24 + (105 – 79)/(210 – 79 – 79) * 6 = 24 + 156/52 = 27. Option A

30. Mean = 90 – 110/50 = 90 – 2.2 = 87.8. Option D

3 MEASURES OF DISPERSION
(Average of Second Order)

THEORY

Introduction:

- Dispersion is defined as deviation or scattering of values from their central values i.e, average (Mean, Median or Mode but preferably Mean or Median)

- Dispersion discovers variability in uniformity.

- In other words, dispersion measures the degree or extent to which the values of a variable deviate from its average

- Dispersion indicates the degree of heterogeneity among observation and as heterogeneity increases dispersion increases

- If all values are equal then any measure of dispersion is always zero

- All measures of dispersion are positive

- All measures of dispersions are independent of the change of origin but dependent on the change of scale

- All pre requisites of a good measure of central tendency are equally applicable for good measure of dispersion

- TWO DISTRIBUTIONS MAY HAVE;
  - i. Same central tendency and same dispersion
  - ii. Different central tendency but same dispersion
  - iii. Same central tendency but different dispersion
  - iv. Different central tendency and different dispersion

Types of Measures of Dispersion

There are two types of measures of dispersion,

Absolute Measure	Relative Measure
a. These measures of dispersion will have the same units as those of the variables	a. These are usually expressed as ratios or percentages and hence unit free
b. Absolute measures are related to the distribution itself.	b. Relative measures are used i) to compare variability between two or more series. ii) To check the relative accuracy of the data

MEASURES OF DISPERSION (AVERAGE OF SECOND ORDER)

A good measure of dispersion should obey conditions similar to those for a satisfactory average and are as follows :

i. It should be rigidly defined.

ii. It should be based on all observations.

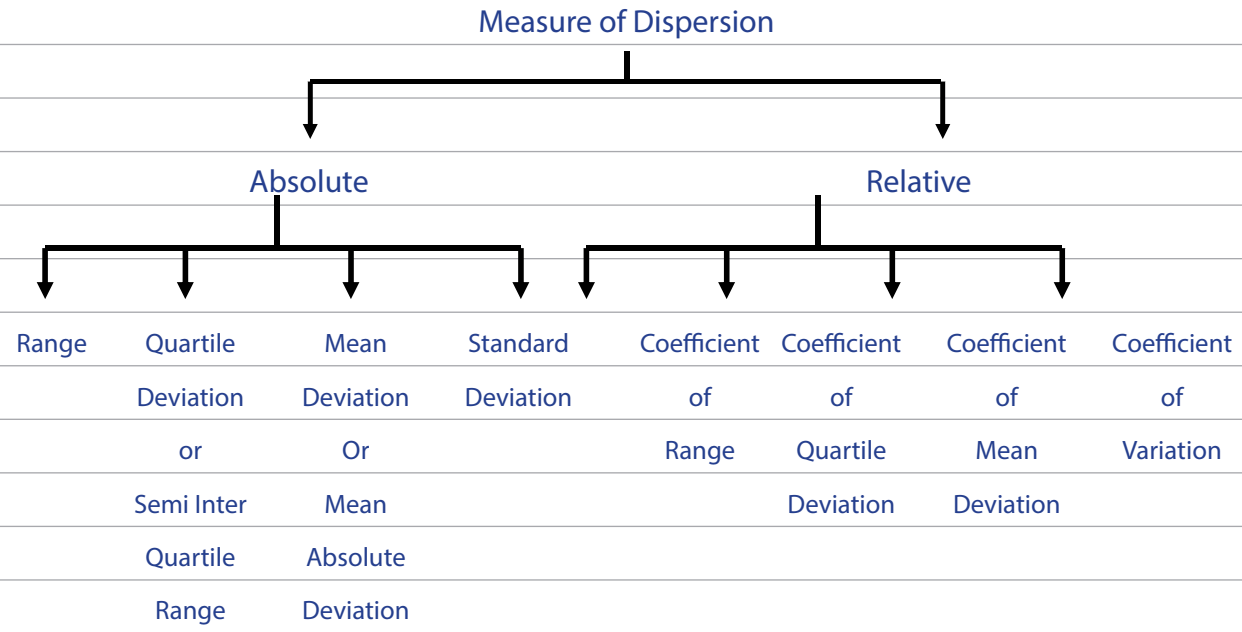
iii. It should be readily comprehensible.

iv. It should be fairly easily calculated.

v. It should affected as little as possible by fluctuations of sampling;

vi. It should readily lend itself to algebraic treatment and

vii. It should be east affected by the presence by extreme values



RANGE

- It is the quickest measure, of finding out Dispersion
- It does not depend on all observations
- It's a crude method of finding out dispersion and most unreliable
- Range is unaffected by the presence of frequency
- Range is independent of the change of origin but dependent on change of scale
- If $y=a\pm bx$
  $R(y)=|b| \times R(x)$

Calculation Of Range:

- For simple series and simple Frequency Distribution :
  Range = Highest Value – Lowest Value (H – L).
- For grouped frequency distribution:
  - Range = Upper boundary of last class – Lower boundary of 1st class
  - Range = Upper Limit of last class – Lower limit of 1st class + 1
- Co-efficient of Range (Relative Range) = $\frac{H - L}{H + L} \times 100$

Quartile Deviation or Semi-inter quartile Range:

- QD is defined as the half of the range between the quartiles
- It is based on the upper and the lower Quartile and covers 50% of the observations.
- It does not depend on all observations
- For distributions with the Open Ends Q.D is the best and only measure of dispersion.
- QD is independent of the change of Origin but dependent on the change of Scale.
- If $y=a\pm bx$
  $QD(y)=|b| \times QD(x)$
- Quartile Deviation (QD) = $\frac{Q_3 - Q_1}{2}$, Where Q3 is the upper quartile and Q1 is the lower quartile.
- Co-efficient of QD(Relative Measure) = $\frac{QD}{Median} \times 100 = \frac{\frac{Q_3 - Q_1}{2}}{Q_2} \times 100 = \frac{Q_3 - Q_1}{2Q_2} \times 100$
- For symmetrical distribution; $Q_2 = \frac{Q_1 + Q_3}{2}$, i.e., median is the average of two extreme quartiles.

  Thus coefficient of QD for symmetrical distribution = $\frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} \times 100 = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

Mean Deviation / Mean Absolute Deviation

- It is based on all observations and hence it provides much better dispersion than Range and Quartile Deviation
- Mean deviation of a set of values of a variable is defined as the AM of the Absolute Deviation taken about Mean, Median or Mode.(Preferably AM or Median)
- Absolute Deviation implies Deviation without any regard to sign
- If nothing is specified Mean Deviation will imply Deviation about AM only.

- Since sum of Deviations is least when Deviations are taken about Median hence MD about Median will have the least value.

- MD is the independent of the change of origin but dependent on the change of scale

- If $y=a\pm bx$

  $MD(y)=|b| \times MD(x)$

- Formula to calculate Mean Deviation:

Simple Series	Simple / Grouped Frequency Distribution
$MD = \frac{\sum\|x - \bar{x}\|}{n}$	$MD = \frac{\sum f\|x - \bar{x}\|}{\sum f}$
$MD = \frac{\sum\|x - M\|}{n}$	$MD = \frac{\sum f\|x - M\|}{\sum f}$

Where n = number of observation

$\sum f=N$ = Total frequency

$\bar{x}$ =A.M

M = Median

X=Either actual values of the variables or mid values if it a group frequency distributions

- Coefficient of MD(Relative Measure) = $\frac{MD}{Mean/Median} \times 100$

Standard Deviation

- It is the best measure and the most commonly used Measure of Dispersion.

- It takes into consideration the magnitude of all the observations and gives the minimum value of dispersion possible.

- SD has all the pre-requisites of a good measure of dispersion, except the fact that it gets unduly affected by the presence of extreme values,

- It is also known as Root Mean Square Deviation about mean.

- It is denoted by σ

- SD^2 = Variance= σ^2
- If all observations are equal variance =SD=0

- SD is the independent of the change of origin but dependent on the change of scale

- If $y=a\pm bx$

  $SD(y)=|b| \times SD(x)$

  $V(y)=b^2 \times v(x)$

Definition of SD:

- SD of a set of values of a variable is defined as the positive Square Root of the AM of the Square of Deviations of the values from their AM

- Thus, SD is also known as Root - Mean – Square - Deviations (RMSD)

Calculation of SD

Simple Series(Without Frequency)	Simple /Grouped Frequency Distribution
i) $\sigma=\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$	i) $\sigma=\sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}}$
ii) $\sigma=\sqrt{\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2}$	ii) $\sigma=\sqrt{\frac{\sum fx^2}{\sum f}-\left(\frac{\sum fx}{\sum f}\right)^2}$
iii) $\sigma_x=\sqrt{\frac{\sum d^2}{n}-\left(\frac{\sum d}{n}\right)^2}\times i$	iii) $\sigma_x=\sqrt{\frac{\sum fd^2}{\sum f}-\left(\frac{\sum fd}{\sum f}\right)^2}\times i$

- Where, $d=\frac{x-A}{i}$,

x= mid-values if it is a grouped frequency distribution or original values if it is a discrete series

A = Assumed Mean i.e., a value arbitrarily chosen from mid-values or any other value.

i = class width or any arbitrary value

Note1 : Use form i) when you find that $\bar{x}$ is whole number

Note2 : Use form ii) when the value of the variable x are small

Note3 : Use Form iii) when you find that the values of x are large $\bar{x}$ is not a whole number(usually to be used for grouped frequency distribution)

USEFUL RESULTS:

- SD of two numbers is the half of their absolute difference(Range), i.e., if numbers are a and b, then SD = $\left|\frac{a-b}{2}\right|$
- Variance of first "n" natural numbers (1, 2, 3, ..., n) is $\frac{n^2-1}{12}$
- Sum of the squares of observations $\sum x^2 = n(\sigma^2 + \bar{x}^2)$

Formula for combined or composite or pooled S.D. of two groups

	Group I	Group II
Numbers	n_1	n_2
Mean	$\overline{x_1}$	$\overline{x_2}$
Standard Deviation	σ_1	σ_2

- Step 1 – Find Combined Mean: $\bar{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2}$
- Step 2 – Find Deviations : $d_1 = \overline{x_1} - \bar{x}$ $\quad d_2 = \overline{x_2} - \bar{x}$
- Step 3 – Use Formula: $\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}$
- Coefficient of Variation (C.V)(Relative Measure) = $\frac{SD}{Mean} \times 100 = \frac{\sigma}{\bar{x}} \times 100$
- C.V is the best relative measure of dispersion
- C.V is used to compare variability or consistency between 2 or more series
- More C.V implies more variability indicating thereby less stability or consistency and vice versa.
- Regarding choice of an item always choose that item which has less C.V, because the item with lower C.V is more stable.

CLASSWORK SECTION

RANGE

1. What is the coefficient of range for the following wages of 8 workers?
₹ 80, ₹ 65, ₹ 90, ₹ 60, ₹ 75, ₹ 70, ₹ 72, ₹ 85.
(a) ₹ 30 (b) ₹ 20 (c) 30 (d) 20

2. If R_x and R_y denote ranges of x and y respectively where x and y are related by 3x+2y+10=0,
what would be the relation between x and y?
(a) $R_x = R_y$ (b) $2R_x = 3R_y$ (c) $3R_x = 2R_y$ (d) $R_x = 2R_y$

3. What is the coefficient of range for the following distribution?

Class Interval :	10-19	20-29	30-39	40-49	50-59
Frequency:	11	25	16	7	3

(a) 22 (b) 50 (c) 72.46 (d) 75.82

4. If the range of x is 2, what would be the range of −3x +50 ?
(a) 2 (b) 6 (c) −6 (d) 44

QUARTILE DEVIATION

5. The quartiles of a variable are 45, 52 and 65 respectively. Its quartile deviation is
(a) 10 (b) 20 (c) 25 (d) 8.30.'

6. If x and y are related as 3x+4y = 20 and the quartile deviation of x is 12, then the quartile deviation of y is
(a) 16 (b) 14 (c) 10 (d) 9

7. Quartile Deviation for the data 1, 3, 4, 5, 6, 6, 10 is
(a) 3 (b) 1 (c) 6 (d) 1.5

8. The value of appropriate measure of dispersion for the following distribution of daily Wages

Wages (₹):	Below 30	30-39	40-49	50-59	60-79	Above 80
No. of workers	5	7	18	32	28	10

is given by

(a) ₹11.03 (b) ₹ 10.50 (c) 11.68 (d) ₹11.68.

MEAN DEVIATION

9. What is the value of mean deviation about mean for the following numbers?
5, 8, 6, 3, 4.

(a) 5.20 (b) 7.20 (c) 1.44 (d) 2.23

10. The coefficient of mean deviation about mean for the first 9 natural numbers is

(a) 200/9 (b) 80 (c) 400/9 (d) 50.

11. If the relation between x and y is 5y−3x = 10 and the mean deviation about mean for x is 12, then the mean deviation of y about mean is

(a) 7.20 (b) 6.80 (c) 20 (d) 18.80.

12. If two variables x and y are related by 2x + 3y −7 =0 and the mean and mean deviation about mean of x are 1 and 0.3 respectively, then the coefficient of mean deviation of y about its mean is

(a) −5 (b) 12 (c) 50 (d) 4.

13. The mean deviation about mode for the numbers 4/11, 6/11, 8/11, 9/11, 12/11, 8/11 is

(a) 1/6 (b) 1/11 (c) 6/11 (d) 5/11.

14. What is the mean deviation about mean for the following distribution?

Variable:	5	10	15	20	25	30
Frequency:	3	4	6	5	3	2

(a) 6.00 (b) 5.93 (c) 6.07 (d) 7.20

15. What is the mean deviation about median for the following data?

X	3	5	7	9	11	13	15
F	2	8	9	16	14	7	4

(a) 2.50 (b) 2.46 (c) 2.43 (d) 2.37

16. What is the coefficient of mean deviation for the following distribution of heights?
Take deviation from AM.

Height in inches:	60-62	63-65	66-68	69-71	72-74
No. of students:	5	22	28	17	3

(a) 2.30 (b) 3.45 (c) 3.82 (d) 2.48

17. The mean deviation of weights about median for the following data:

Weight (lb) :	131-140	141-150	151-160	161-170	171-180	181-190
No. of persons :	3	8	13	15	6	5

Is given by

(a) 10.97 (b) 8.23 (c) 9.63 (d) 11.45.

STANDARD DEVIATION

18. What is the standard deviation of 5, 5, 9, 9, 9, 10, 5, 10, 10?

(a) $\sqrt{14}$ (b) $\frac{\sqrt{42}}{3}$ (c) 4.50 (d) 8

19. If the mean and SD of x are a and b respectively, then the SD of $\frac{(x-a)}{b}$ is

(a) −1 (b) 1 (c) ab (d) a/b.

20. What is the coefficient of variation of the following numbers?
53, 52, 61, 60, 64.

(a) 8.09 (b) 18.08 (c) 20.23 (d) 20.45

21. If the SD of x is 3, what is the variance of (5−2x)?

(a) 36 (b) 6 (c) 1 (d) 9

22. If x and y are related by 2x+3y+4 = 0 and SD of x is 6, then SD of y is

(a) 22 (b) 4 (c) $\sqrt{5}$ (d) 9

23. If the SD of the 1st n natural numbers is 2, then the value of n must be

(a) 2 (b) 7 (c) 6 (d) 5

24. If x and y are related by y = 2x+ 5 and the SD and AM of x are known to be 5 and 10 respectively, then the coefficient of variation of y is

(a) 25 (b) 30 (c) 40 (d) 20

25. The mean and SD for a, b and 2 are 3 and $\frac{2}{\sqrt{3}}$ respectively, The value of ab would be

(a) 5 (b) 6 (c) 11 (d) 3

26. What is the standard deviation from the following data relating to the age distribution of200 persons?

Age (year) :	20	30	40	50	60	70	80
No. of people:	13	28	31	46	39	23	20

(a) 15.29 (b) 16.87 (c) 18.00 (d) 17.52

27. What is the coefficient of variation for the following distribution of wages?

Daily Wages (₹):	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
No. of workers	17	28	21	15	13	6

(a) ₹ 14.73 (b) 14.73 (c) 26.93 (d) 20.82

28. Which of the following companies A and B is more consistent so far as the payment of dividend is concerned ?

Dividend paid by A :	5	9	6	12	15	10	8	10
Dividend paid by B :	4	8	7	15	18	9	6	6

(a) A (b) B

(c) Both (a) and (b) (d) Neither (a) nor (b)

COMBINED STANDARD DEVIATION

29. If two samples of sizes 30 and 20 have means as 55 and 60 and variances as 16 and 25 respectively, then what would be the SD of the combined sample of size 50?

(a) 5.00 (b) 5.06 (c) 5.23 (d) 5.35

CORRECTION IN STANDARD DEVIATION

30. The mean and SD of a sample of 100 observations were calculated as 40 and 5.1 respectively by a CA student who took one of the observations as 50 instead of 40 by mistake. The correct value of SD would be

(a) 4.90 (b) 5.00 (c) 4.88 (d) 4.85.

THEORETICAL ASPECTS

31. Which of the following statements is correct?

(a) Two distributions may have identical measures of central tendency and dispersion.

(b) Two distributions may have the identical measures of central tendency but different measures of dispersion.

(c) Two distributions may have the different measures of central tendency but identical measures of dispersion.

(d) All the statements (a), (b) and (c).

32. Dispersion measures

(a) The scatterness of a set of observations

(b) The concentration of a set of observations

(c) Both (a) and (b)

(d) Neither (a) and (b).

33. When it comes to comparing two or more distributions we consider

(a) Absolute measures of dispersion (b) Relative measures of dispersion

(c) Both (a) and (b) (d) Either (a) or (b).

34. Which one is easier to compute?

(a) Relative measures of dispersion (b) Absolute measures of dispersion

(c) Both (a) and (b) (d) Range

35. Which one is an absolute measure of dispersion?

(a) Range (b) Mean Deviation

(c) Standard Deviation (d) All these measures

36. Which measure of dispersion is most usefull?

(a) Standard deviation (b) Quartile deviation
(c) Mean deviation (d) Range

37. Which measures of dispersions is not affected by the presence of extreme observations?

(a) Range (b) Mean deviation
(c) Standard deviation (d) Quartile deviation

38. Which measure of dispersion is based on the absolute deviations only?

(a) Standard deviation (b) Mean deviation
(c) Quartile deviation (d) Range

39. Which measure is based on only the central fifty percent of the observations?

(a) Standard deviation (b) Quartile deviation
(c) Mean deviation (d) All these measures

40. Which measure of dispersion is based on all the observations?

(a) Mean deviation (b) Standard deviation
(c) Quartile deviation (d) (a) and (b) but not (c)

41. The appropriate measure of dispersion for open-end classification is

(a) Standard deviation (b) Mean deviation
(c) Quartile deviation (d) All these measures.

42. The most commonly used measure of dispersion is

(a) Range (b) Standard deviation
(c) Coefficient of variation (d) Quartile deviation.

43. Which measure of dispersion has some desirable mathematical properties?

(a) Standard deviation (b) Mean deviation
(c) Quartile deviation (d) All these measures

44. If the profits of a company remains the same for the last ten months, then the standard deviation of profits for these ten months would be ?

(a) Positive (b) Negative (c) Zero (d) (a) or (c)

45. Which measure of dispersion is considered for finding a pooled measure of dispersion after combining several groups?

(a) Mean deviation (b) Standard deviation
(c) Quartile deviation (d) Any of these

46. A shift of origin has no impact on

(a) Range (b) Mean deviation
(c) Standard deviation (d) All these and quartile deviation.

47. The range of 15, 12, 10, 9, 17, 20 is

(a) 5 (b) 12 (c) 13 (d) 11.

48. The standard deviation of 10, 16, 10, 16, 10, 10, 16, 16 is

(a) 4 (b) 6 (c) 3 (d) 0.

49. For any two numbers SD is always

(a) Twice the range (b) Half of the range
(c) Square of the range (d) None of these.

50. If all the observations are increased by 10, then

(a) SD would be increased by 10
(b) Mean deviation would be increased by 10
(c) Quartile deviation would be increased by 10
(d) All these three remain unchanged.

51. If all the observations are multiplied by 2, then

(a) New SD would be also multiplied by 2
(b) New SD would be half of the previous SD
(c) New SD would be increased by 2
(d) New SD would be decreased by 2.

52. "Root –Mean Square Deviation from Mean" is

(a) Standard deviation (b) Quartile deviation
(c) both (d) none

53. The amount of variation is designated as ——— measure of dispersion.

(a) relative (b) absolute (c) both (d) none

54. The degree of variation is designated as ——— measure of dispersion.

(a) relative (b) absolute (c) both (d) none

ANSWERS - SUMS				ANSWERS - THEORITICAL ASPECTS			
Q. No.	Ans	Q. No.	Ans	Q. No.	Ans	Q. No.	Ans
1	d	21	a	31	d	51	a
2	c	22	b	32	a	52	a
3	c	23	b	33	b	53	b
4	b	24	c	34	d	54	a
5	a	25	c	35	d		
6	d	26	b	36	a		
7	d	27	b	37	d		
8	a	28	a	38	b		
9	c	29	b	39	b		
10	c	30	b	40	d		
11	a			41	c		
12	b			42	b		
13	a			43	a		
14	c			44	c		
15	d			45	b		
16	b			46	d		
17	a			47	d		
18	b			48	c		
19	b			49	b		
20	a			50	d		

HOMEWORK SECTION

1. A student obtained the mean and standard deviation of 100 observations as 40 and 5.1 respectively. It was later discovered that he had wrongly copied down an observation as 50 instead of 40. The correct standard deviation is:
   (a) 5 (b) 6 (c) 3 (d) 7

2. For a Symmetrical distribution, quartile deviation and the standard deviation are related by
   (a) 5 Q.D. = 4 S.D. (b) 4 Q.D. = 5 S.D.
   (c) 2 Q.D. = 3 S.D. (d) 3 Q.D. = 2 S.D.

3. If two samples of sizes 30 and 20 have means as 55 and 60 and variances as 16 and 25 respectively, then what would be the S.D. of the combined sample size 50?
   (a) 5.33 (b) 5.17 (c) 5.06 (d) 5

4. If two variables x and y are related by $2x + 3y - 7 = 0$ and the mean and mean deviation about mean of x are 1 and 0.3 respectively, then the co-efficient of mean deviation of y about mean is
   (a) −5 (b) 4 (c) 12 (d) 50

5. Measures of dispersion are called average of the ____ order.
   (a) 1st (b) 2nd (c) 3rd (d) none

6. In a set of 100 observations, taking assumed mean as 4, the sum of the deviations is −11 cm, and the sum of the squares of these deviations is 257 cm^2. The coefficient of variation is
   (a) 41.13% (b) 42.13% (c) 40.13% (d) none

7. _____ are used for measuring central tendency, dispersion and skewness:
   (a) Median (b) Deciles
   (c) Percentiles (d) Quartiles

8. Which of the following companies A or B is more consistent so far as the payment of dividend is concerned?

Dividend paid by A : 5	9	6	12	15	10	8	10
Dividend paid by B : 4	8	7	15	18	9	6	6

(a) A (b) B
(c) Both A & B (d) Neither A nor B

9. What is the coefficient of range for the following distribution?

Class Interval	10–19	20–29	30–39	40–49	50–59
Frequency	11	25	16	7	3

(a) 22 (b) 50 (c) 75.82 (d) 72.46

10. A sample of 35 observations has the mean 80 and S.D. as 4. A second sample of 65 observations from the same population has mean 70 and S.D. 3. The S.D. of the combined sample is:

(a) 5.85 (b) 5.58 (c) 10.23 (d) None of these

11. If x and y are related as $3x - 4y = 20$ and the quartile deviation of x is 12, then the quartile deviation of y is :

(a) 14 (b) 15 (c) 16 (d) 9

12. The best measure of dispersion is :

(a) Q.D. (B) M.D. (c) Range (d) S.D.

13. If the mean and S.D. of x and a and b respectively, then the S.D. of $\frac{x-a}{b}$ is :

(a) a/b (b) −1 (c) 1 (d) ab

14. Suppose a population A has 100 observations 101, 102, 103, . . . 200 and another population B has 100 observations 151, 152, 153, . . . 250. If V_A and V_B represents the variance of the two populations respectively, then V_A / V_B =

(a) 9/4 (b) 1 (c) 4/9 (d) 2/3

15. The mean and S.D. for group of 100 observations are 65 and 7.03 respectively. If 60 of these observations have mean and S.D. as 70 and 3 respectively, what is the S.D. for the group comprising 40 observations?

(a) 2.03 (b) 4.03 (c) 8.03 (d) 9.33

16. The quartile deviation for the data is:

x	2	3	4	5	6
f	3	4	8	4	1

(a) 1/4 (b) 1/2 (c) 0.875 (d) 1

17. If X and Y are two independent random variables then v(x + y) is:

(a) v(x) + v(y) (b) v(x) + v(y) − 2v(x, y)

(c) v(x) + v(y) + 2v(x, y) (d) v(x) − v(y)

18. Mean and S.D. of x is 50 and 5 respectively. Find mean and S.D. of $\frac{x - 50}{5}$.

(a) (1, 0) (b) (0, 1)

(c) (1, 1) (d) (0, −1)

19. Mean and S.D. of a given set of observations is 1,500 and 400 respectively. If there is an increment of 100 in the first year and each observation is hiked by 20% in 2nd years, then find new mean and S.D.

(a) 1920, 480 (b) 1920, 580

(c) 1600, 480 (d) 1600, 400

20. If 5 is subtracted from each observation of some certain item then its co-efficient of variation is 10% and if 5 is added to each item then its coefficient of variation is 6%. Find original coefficient of variation.

(a) 8% (b) 7.5%

(c) 4% (d) none of these

21. Inter Quartile Range is _____ of Quartile Deviation.

(a) Half (b) Double

(c) Triple (d) Equal

22. The sum of squares of deviation from mean of 10 observations is 250. Mean of the data is 10. Find the co-efficient of variation.

(a) 10% (b) 25%

(c) 50% (d) 0%

23. If L_1 = highest observation and L2 = smallest observation, then Coefficient of Range

(a) $\frac{L_1 \times L_2}{L_1 / L_2} \times 100$ (b) $\frac{L_1 - L_2}{L_1 + L_2} \times 100$

(c) $\frac{L_1 + L_2}{L_1 - L_2} \times 100$ (d) $\frac{L_1 / L_2}{L_1 \times L_2} \times 100$

24. The equation of a line is 5x + 2y = 17. Mean deviation of y about mean is 5. Calculate mean deviation of x about mean

(a) −2 (b) 2 (c) −4 (d) none

25. If variance of x is 5, then find the variance of (2 − 3x)

(a) 10 (b) 45 (c) 5 (d) −13

26. The variance of data : 3, 4, 5, 8 is

(a) 4.5 (b) 3.5 (c) 5.5 (d) 6.5

27. Given the observations: 4, 9, 11, 14, 37. The Mean deviation about the median is

(a) 11 (b) 8.5 (c) 7.6 (d) 7.45

28. If all observations in a distribution are increased by 6, then the variance of the series will be _____.

(a) Increased (b) Decreased

(c) Unchanged (d) None of these

29. The standard deviation of the weights (in kg) of the students of a class of 50 students was calculated to be 4.5 kg. Later on it was found that due to some fault in weighing machine, the weight of each student was under measured by 0.5 kg. The correct standard deviation of the weight will be:

(a) Less than 4.5 (b) Greater than 4.5

(c) Equal to 4.5 (d) Can not be determined

30. For Normal distribution he relation between quartile deviation (Q.D.) and standard deviation (S.D.) is

(a) Q.D. > S.D. (b) Q.D. < S.D.

(c) Q.D. = S.D. (d) None of the above

31. If standard deviation of first 'n' natural numbers is 2 then value of 'n' is

(a) 10 (b) 7 (c) 6 (d) 5

32. The standard deviation is independent of change of

(a) Scale (b) Origin

(c) Both origin and scale (d) None of these

33 In a normal distribution, the relationship between the three most commonly used measures of dispersion are:

(a) Standard Deviation > Mean Deviation > Quartile Deviation

(b) Mean Deviation > Standard Deviation > Quartile Deviation

(c) Standard Deviation > Quartile Deviation > Mean Deviation

(d) Quartile Deviation > Mean Deviation > Standard Deviation

34. If Standard deviation of x is σ, then Standard deviation of $\frac{ax+b}{c}$, where a, b and c (c ± 0) are arbitrary constants, will be

(a) σ (b) $\frac{a\sigma+b}{c}$ (c) $\frac{a}{c}.\sigma$ (d) $\left|\frac{a}{c}\right|\sigma$

35. Which of the following measures of dispersion is used for calculating the consistency between two series?

(a) Quartile deviation (b) Standard deviation

(c) Coefficient of variation (d) None of the above

36. If sum of squares of the values = 3390, n = 30 and standard deviation = 7, find out the mean.

(a) 113 (b) 210 (c) 8 (d) None of these

37. If the mean of a frequency distribution is 100 and coefficient of variation is 45% then standard deviation is:

(a) 45 (b) 0.45 (c) 4.5 (d) 450

38. Find at the variance given that the Arithmetic Mean = (8 + 4)/2

(a) 2 (b) 6 (c) 1 (d) 4

39. Coefficient of mean deviation about mean for the first 9 natural numbers is

(a) 200/9 (b) 80 (c) 400/9 (d) 50

40. If mean = 5, Standard deviation = 2.6, median = 5 and quartile deviation = 1.5, then the coefficient of quartile deviation equals

(a) 35 (b) 39 (c) 30 (d) 32

41. What will be the probable value of mean deviation? When $Q_3 = 40$ and $Q_1 = 15$.

(a) 17.50 (b) 18.75 (c) 15.00 (d) None of the above

42. The formula for range of middle 50% items of a series is:

(a) $Q_3 - Q_1$ (b) $Q_3 - Q_2$ (c) $Q_2 - Q_1$ (d) $\frac{Q_3 - Q_1}{2}$

43. If the first quartile is 142 and semi-inter quartile range is 18, then the value of median is:

(a) 151 (b) 160 (c) 178 (d) None of these

44. The quartile deviation is:

(a) 2/3 of S.D.
(b) 4/5 of S.D.
(c) 5/6 of S.D.
(d) None of these

45. The standard deviation of a variable x is known to be 10. The standard deviation of 50 + 5x is

(a) 50 (b) 100 (c) 10 (d) 500

46. Coefficient f quartile deviation is equal to

(a) Quartile deviation × 100/median
(b) Quartile deviation × 100/mean
(c) Quartile deviation × 100/mode
(d) none

47. If all the observations are increased by 5, then

(a) S.D. would be increased by 5
(b) Mean deviation would be increased by 5
(c) Quartile deviation would be increased by 5
(d) All the three would not be increased by 5

48. What is value of mean deviation about mean from the number 5, 8, 6, 3 and 4?

(a) 5.20 (b) 7.20 (c) 1.44 (d) 2.23

49. For the observation of 6, 4, 1, 6, 5, 10, 4, 8 the range is:

(a) 10 (b) 9 (c) 8 (d) None

50. If a variance of a random variable 'x' is 23, then what is variance of y= 2x + 10?

(a) 56 (b) 33 (c) 46 (d) 92

51. If variance = 148.6 and $\bar{x}$, =40 then the coefficient of variation is:

(a) 37.15

(b) 30.48

(c) 33.75

(d) None of the above

52. The SD of first n natural number is _____

(a) $\sqrt{\frac{n^2-1}{12}}$

(b) $\sqrt{\frac{n(n+1)}{12}}$

(c) $\sqrt{\frac{n(n-1)}{6}}$

(d) None of these

53. If mean and coefficient of variation of the marks of 10 students is 20 and 80 respectively. What will be variance of them?

(a) 256 (b) 16 (c) 25 (d) None of these

54. If same amount is added to or subtracted from all the values of an individual series then the standard deviation and variance both shall be _____

(a) changed

(b) unchanged

(c) same

(d) none of these

55. If arithmetic mean and coefficient of variation of x are 10 and 40, respectively then the variance of $y = -15 + \frac{3x}{2}$ will be:

(a) 64 (b) 81 (c) 49 (d) 36

56. Mean deviation is the least when deviations are taken from

(a) Mean (b) Median (c) Mode (d) Harmonic mean

57. $\frac{(Q_3 - Q_1)}{(Q_3 + Q_1)}$ is known as

(a) Coefficient of Range
(b) Coefficient of Q.D.
(c) Coefficient of S.D.
(d) Coefficient of M.D.

58. If the S.D. of the 1st n natural nos. is $\sqrt{30}$. Then the value of n is

(a) 19 (b) 20 (c) 21 (d) None

59. If the range of a set of values 65 and maximum value in the set is 83, then the minimum value in the set is

(a) 74 (b) 9 (c) 18 (d) None of the above

60. If the variance of 5, 7, 9 and 11 is 4, then the coefficient of variation is:

(a) 15 (b) 25 (c) 17 (d) 19

61. Standard Deviation for the marks obtained by a student in monthly test in mathematic (out of 50) as 30, 35, 25, 20, 15 is

(a) 25 (b) $\sqrt{50}$ (c) $\sqrt{30}$ (d) 50

62. If the standard deviation for the marks obtained by a student in monthly test is 36, then the variance is

(a) 6 (b) 36 (c) 1296 (d) None of the above

ANSWER

1	a	21	b	41	c	61	b
2	d	22	c	42	a	62	c
3	c	23	b	43	b		
4	c	24	b	44	a		
5	b	25	b	45	a		
6	a	26	b	46	a		
7	d	27	c	47	d		
8	a	28	c	48	c		
9	d	29	c	49	b		
10	a	30	b	50	d		
11	d	31	b	51	b		
12	d	32	b	52	a		
13	c	33	a	53	a		
14	b	34	d	54	b		
15	b	35	c	55	d		
16	d	36	c	56	b		
17	a	37	a	57	b		
18	b	38	d	58	a		
19	a	39	c	59	c		
20	b	40	c	60	b		

HOMEWORK SOLUTION

1. (a) Correct $\sum x = (100 * 40) - 50 + 40 = 3990$
Correct $\sum x^2 = 100[(5.1)^2 + (40)^2] - 50^2 + 40^2 = 161701$
Correct $SD^2 = 161701/100 - (3990/100)^2 = 25$
Correct $SD = 5$

2. (d) For a moderately skewed distribution: SD = 1.5 times the QD.

3. (c) Combined Mean = $(30*55 + 20*60)/(30+20) = 57$
$D1 = 55 - 57 = -2$
$D2 = 60 - 57 = 3$
Combined $SD^2 = [30(16 + 4) + 20(25 + 9)]/50 = 1280/50 = 25.6$
Combined $SD = \sqrt{25.6} = 5.06$

4. (c) Mean(y) = $[7 - 2(1)]/3 = 5/3$
MD (y) = $2(0.3)/3 = 0.2$
Coefficient of MD (y) = $0.2 / 5/3 = 0.6/5 = 0.12 * 100 = 12$

5. (b) Measures of dispersion are known as averages of the second order.

6. (a) Mean = $4 - 11/100 = 3.89$
$\sum x^2 = 257 + 8(389) - 16(100) = 1769$
$SD^2 = 1769/100 - (3.89)^2 = 2.5579$
$SD = 1.60$
Coefficient of variation = $1.60/3.89 * 100 = 41.13\%$

7. (d) Quartiles are used for measuring central tendency, dispersion and skewness.

8. (a) Company A:

Mean = (5+9+6+12+15+10+8+10)/8 = 9.375

$\sum x^2$ = 25+81+36+144+225+100+64+100 = 775

SD^2 = 775/8 – $(9.375)^2$ = 8.984375

SD = 3

COV = 3/9.375 * 100 = 32%.

Company B:

Mean = (4+8+7+15+18+9+6+6)/8 = 9.125

$\sum x^2$ = 16+64+49+225+324+81+36+36 = 831

SD^2 = 831/8 – (9.125)2 = 20.609375

SD = 4.54

COV = 4.54/9.125 * 100 = 49.75%

Company A is more consistent. [As COV(A) < COV(B)]

9. (d) Highest Mark = 59.5

Lowest Mark = 9.5

Coefficient of Range = (H – L)/(H + L) * 100 = (59.5 – 9.5)/(59.5 + 9.5) * 100 = 50/69 = 72.46

10. (a) Combined Mean = (35*80 + 65*70)/(35+65) = 73.50

D1 = 80 – 73.50 = 6.50

D2 = 70 – 73.50 = – 3.50

Combined SD2 = [35(16 + 42.25) + 65(9 + 12.25)]/100 = (2038.75+1381.25)/100 = 34.20

Combined SD = $\sqrt{34.20}$ = 5.85

11. (d) QD(y) = 3(12)/4 = 9,

12. (d) The best measure of dispersion is Standard Deviation.

13. (c) Y = (X – a)/b

b.SD(Y) = SD(X) = b

SD(Y) = b/b = 1

14. (b) $A = B - 50$

$V(A) = V(B)$

$V(A) / V(B) = 1$

15. (b) $65*100 = (60 * 70) + (40 * X)$

$6500 = 4200 + 40X$

X (Mean of the second group) $= 57.5$

$D1 = 70 - 65 = 5$

$D2 = 57.5 - 65 = -7.5$

$100(7.03)^2 = 60(9 + 25) + 40(S^2 + 56.25)$

$(4942.09 - 2040)/40 - 56.25 = S^2 = 16.30225$

$S = 4.03$

16. (d) $QD = \frac{Q_3 - Q_1}{2}$

On solving we get , $Q_1 = 3$ and $Q_3 = 5$

$QD = \frac{5-3}{2}$

$= \frac{2}{2}$

$= 1$

17. (a) $V(X + Y) = V(X) + V(Y)$

18. (b) $Y = (x - 50)/5$

$5Y = x - 50$

$5\ SD(Y) = SD(x) = 5$

$SD(Y) = 5/5 = 1$

$Mean(Y) = (50 - 50)/5 = 0$

19. (a) $Y = 1.2(X + 100) = 1.2X + 120$

$Mean(Y) = 1.2(1500) + 120 = 1920$

$SD(Y) = 1.2\ SD(X) = 1.2 * 400 = 480$

20. (b) $Y = X - 5$

$10 = S/(X' - 5) * 100$

$X' - 5 = 10S$

$Z = X + 5$

$6 = S/(X' + 5) * 100$

$6X' + 30 = 100S$

$10X' - 50 = 100S$

$4X' = 80$

Mean(X) = 20

SD(X) = 1.5

Original COV = 1.5/20 * 100 = 7.5%

21. (b) Inter-Quartile Range is Double of Quartile Deviation.

22. (c) $SD^2 = 250/10 = 25$; SD = 5

COV = 5/10 * 100 = 50%

23. (b) Coefficient of Range = (L1 – L2)/(L1 + L2) * 100.

24. (b) 5 MD(x) = 2 MD(y) = 2*5 = 10

MD(x) = 10/5 = 2.

25. (b) Vx = 5

Vy= ?

Y = 2 -3x

$Vy = b^2 \times Vx$

$= (-3)^2 \times 5$

$= 9 \times 5$

$= 45$

26. (b) Mean = (3 + 4 + 5 + 8) = 20/4 = 5

SD2 = (4 + 1 + 0 + 9)/4 = 14/4 = 3.5

27. (c) Median for the data 4,9,11,14,37

Median 4, 9, 11, 14, 37

M = 11

MD from M = $\frac{\Sigma |x - M|}{n}$

= $\frac{38}{5}$

= 7.6

28. (c) Variance of the series is independent of change of the origin.

29. (c) SD is independent of the change of origin. No change in the SD. SD = 4.5.

30. (b) SD > QD. Or, QD < SD.

31. (b) $S^2 = (n^2 - 1)/12$

$4*12 + 1 = n^2 = 49$

n = 7

32. (b) SD is independent of the change of Origin.

33. (a) In normal distribution SD > MD > QD.

34. (d) y = (ax+b)/c

y = a/c x + b/c

$\sigma y = |b| \sigma x$

$\sigma y = \left|\frac{a}{c}\right| \sigma x$

$\sigma y = \left|\frac{a}{c}\right| \sigma$

35. (c) Standard deviation is used to measure the consistency between two series using Coefficient of variation.

36. (c) $Mean^2 = 3390/30 - SD^2 = 113 - 49 = 64$

Mean = 8.

37. (a) 0.45 = SD/100

SD = 45

38. (d) $\bar{x} = \frac{8+4}{2}$ = (8+4)/2 (given)

$\therefore$ The two numbers are 8 and 4

$\sigma = \frac{Range}{2} = \frac{8-4}{2} = \frac{4}{2} = 2$

Variance = σ^2 = 4

39. (c) Coefficient MD = $\frac{\Sigma|x-\bar{x}|}{n} / \bar{x}$ x 100

$\bar{x} = \frac{n+1}{2} = \frac{9+1}{2}$ = 5 (Mean for n natural no's)

$\Sigma(|x - \bar{x}|)$ = 20

Coefficient of MD = $\frac{\frac{20}{9}}{5}$ x 100

Coefficient of MD = $\frac{400}{9}$

40. (c) Coefficient of QD = QD/Median * 100 = 1.5/5 * 100 = 30

41. (c) QD = (40 − 15)/2 = 12.5

MD = (12.5)*6/5 = 15

42. (a) Range of middle 50% of items = $Q_3 - Q_1$ and it is called inter quartile range.

Whereas $\frac{Q_3 - Q_1}{2}$ is called semi inter quartile range which is quartile deviation

43. (b) 18 = (Q3 − 142)/2

Q3 = 178

Q2 = (178 + 142)/2 = 160

44. (a) The QD = 2/3 SD.

45. (a) SD(50 + 5x) = 5.SD(x) = 5*10 = 50.

46. (a) Coefficient of QD = QD/Median * 100.

47. (d) If all the observations are increased by 5, there is no change in any measure of dispersion, as all measures of dispersion are independent of change of origin.

48. (c) Mean = (5+8+6+3+4)/5 = 26/5 = 5.2
|x − Mean| = 0.2 + 2.8 + 0.8 + 2.2 + 1.2 = 7.2
MD = 7.2/5 = 1.44.

49. (b) Range = 10 − 1 = 9.

50. (d) V(2x + 10) = 4*V(x) = 4*23 = 92.

51. (b) Variance = 148.6, SD = 12.2
Coefficient of variation = 12.2/40 * 100 = 30.48

52. (a) $SD^2 = (n^2 - 1)/12$

53. (a) 80/100 = SD/20; SD = 16; Variance = 16^2 = 256.

54. (b) Both SD and variance remain unchanged.

55. (d) 40/100 = SD/10; SD = 4; Variance = 16
V(−15 + 3x/2) = 9/4 * V(x) = 9/4 * 16 = 36

56. (b) Mead deviation is least when deviations are taken from Median.

57. (b) Coefficient of QD = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

58. (a) 30 = (n2 − 1)/12
n^2 = 30*12 + 1 = 361; n = $\sqrt{361}$ = 19.

59. (c) 65 = 83 − Minimum value.
Minimum value = 83 − 65 = 18.

60. (b) Mean = (5 + 7 + 9 + 11)/4 = 32/4 = 8
SD = $\sqrt{4}$ = 2
COV = 2/8 * 100 = 25%

61. (b) Mean = (30 + 35 + 25 + 20 + 15)/5 = 25

$SD^2 = [(30 - 25)^2 + (35 - 25)^2 + (25 - 25)^2 + (20 - 25)^2 + (15 - 25)^2] / 5$

$= (25 + 100 + 0 + 25 + 100) / 5 = 50$

$SD = \sqrt{50}$.

62. (c) Variance = $36^2 = 1296$.

SELF ASSESSMENT TEST 3
30 Marks

1. What does measure of dispersion indicate:
   a) Correlation between data
   b) Regression between variables
   c) Variability of data
   d) Central tendency of data

2. Standard deviation of two numbers A and B is:
   a) |A − B|
   b) (A − B)/2
   c) |A − B| / 2
   d) (A − B)

3. A shift of origin has no impact on:
   a) Range
   b) Quartile Deviation
   c) Mean Deviation
   d) All of the above

4. Which measures of dispersion is not affected by the presence of extreme values?
   a) Standard Deviation
   b) Range
   c) Mean Deviation
   d) Semi-Inter Quartile Range

5. Which empirical relation is CORRECT?
   a) MD = 8 SD
   b) 8 MD = SD
   c) MD = 0.8 SD
   d) MD = 1.25 SD

6. We shall compute the following to study the deviations of middle 40% portion of a series:
   a) $Q_2 - Q_1$
   b) $Q_3 - Q_1$
   c) $P_{70} - P_{30}$
   d) $D_6 - D_2$

7. Which measure of dispersion is useful in open-end classes?
   a) Range
   b) Mean Deviation
   c) Standard Deviation
   d) Semi-Inter Quartile Range

8. The square of the standard deviation is known as:
   a) Variation
   b) Coefficient of Variation
   c) Coefficient of Standard Deviation
   d) None of the above

9. What empirical relation would you expect to exist between the semi-inter quartile range and the mean deviation for bell shaped distribution which are moderately skewed?

a) Semi IQR = 5/6 * MD
b) Semi IQR = 6/5 * MD
c) 5/6 * Semi IQR = MD
d) None of the above

10. If the profits of a firm remain the same for the last 10 months, then the standard deviation of profits for these 10 months would be?

a) Positive
b) Zero
c) Negative
d) Can't be determined

11. The range find its wide application in:

a) Econometrics
b) Quantum Statistics
c) Statistical Quality Control
d) Psychoroetrics

12. Which measure of dispersion is based on the absolute deviations only?

a) Standard Deviation
b) Range
c) Quartile Deviation
d) Mean Deviation

13. Which measure of dispersion is considered for computing a pooled measure of dispersion after combining several groups?

a) Mean Deviation & Standard Deviation
b) Mean Deviation
c) Standard Deviation
d) Range

14. To compare the variability between two series which also differ on their unit of measurements, the measure usually used is:

a) Standard Deviation
b) Mean Absolute Deviation
c) Coefficient of Variation
d) Inter-Quartile Range

15. Find the standard deviation of 4, 8, 10, 12, 16.

a) 2 b) 4 c) 8 d) None of the above

16. Calculate the variance of 1, 5, 6.

a) 2.16 b) 4.67 c) 6.47 d) 5.47

17. Find the Mean-Deviation about the Arithmetic Mean of the numbers 31, 35, 29, 63, 55, 72, 37.

a) 12.86 b) 14.86 c) 13.78 d) 13.86

18. Find the coefficient of mean deviation of the following numbers: 46, 79, 26, 85, 39, 65, 99, 29, 56, 72.

a) 33.27% b) 34.87% c) 34.23% d) 32.43%

19. The arithmetic mean is 25 and standard deviation is 6.25. Find the coefficient of variation.

a) 20 b) 25 c) 30 d) 50

20. The mean and SD of 20 items were found to be 12 and 6 respectively. On checking it was dis-covered that items which should correctly read as 11 and 21 had been wrongly taken as 15 and 27 respectively. Find the correct SD.

a) 3.5 b) 4.3 c) 5.3 d) 4.5

21. Find mean, if co-efficient of variation is 5% and variance is 4.

a) 30 b) 40 c) 45 d) None of the above

22. The mean of 5 observations is 4.4 and the variance is 8.24. If three of the five observations are 1, 2 and 6, find the other two.

a) 4, 8 b) 9, 4 c) 8, 6 d) 9, 5

23. If in the distribution, $n = 10$, $\sum x = 20$, $\sum x^2 = 200$, then find the value of SD.

a) 2 b) 4 c) 3.5 d) 4.5

24. The two variables X and Y are related by $Y = 10 - 3X$. If the SD of X is 4, what is the SD of Y.

a) 2 b) 12 c) -2 d) None of the above

25. Find the Quartile Deviation of the following data: 12, 10, 17, 14, 19, 21, 27, 30, 32, 28, 34.

a) 8 b) 3.8 c) 8.3 d) 9

26. The standard deviation of 10, 16, 10, 16, 10, 10, 16, 16 is:

a) 6 b) 4 c) 3 d) 9

27. The maximum and minimum values of a series are 60 and 0 respectively. What is the coeffi-cient of range?

a) 1 b) – 1 c) 0 d) None of the above

28. Find the first quartile, if coefficient of quartile deviation = 0.37 and the third quartile is 46.25.

a) 46.62 b) 21.26 c) 21.07 d) 27.08

29. The mean of 200 items is 48 and standard deviation is 3. What is the sum of the squares of these items?

a) 462600 b) 400000 c) 460000 d) None of the above

30. The means of two samples of sizes 50 and 100 respectively are 54.4 and 50.3 and the standard deviation is 8 and 7. Obtain the standard deviation of the sample size 150 obtained by combining the two samples.

a) 5.79 b) 6.79 c) 7.59 d) 7.37

EXPLANATORY ANSWERS

1. Measures of dispersion indicate the variability of data. Option C

2. SD of two numbers is average if their absolute range = ½ |A − B|. Option C

3. All measures of dispersion are independent of change/shift of origin. Option D

4. Quartile deviation or Semi-Inter Quartile Range is not affected by the presence of extreme val-ues. Option D

5. Mean Deviation = 80% of Standard Deviation. Option C

6. Middle 40% means to the left and right, we need to leave 60% data, 30% on both sides. Thus middle 40% is correctly represented by $P_{70} - P_{30}$. Option C

7. Quartile deviation best fits in case of open end classes. Option D

8. Square of SD = Variance. Option D

9. Quartile Deviation = 5/6 Mean Deviation. Option A

10. If all values are same, SD = 0. Option B

11. Range is widely used in statistical quality control measures. Option C

12. Mean deviation is based on absolute deviations. Option D

13. Standard deviation is used to measure pooled measure of dispersion. Option C

14. With different units of measurements, co-efficient of variation is the best relative measure of dispersion. Option C

15. Mean = (4 + 8 + 10 + 12 + 16)/5 = 50/5 = 10
SD^2 = [(4-10)^2+(8-10)^2+(10-10)^2+(12-10)^2+(16-10)^2]/5 = (36+4+0+4+36)/5 = 16
SD = 4. Option B

16. Mean = (1 + 5 + 6)/3 = 12/3 = 4
Variance= [(1-4)^2 + (5-4)^2 + (6-4)^2]/3 = (9 + 1 + 4)/3 = 14/3 = 4.67. Option B

17. AM = (31+35+29+63+55+72+37)/7 = 46
MD = (15 + 11 + 17 + 17 + 9 + 26 + 9)/7 = 104/7 = 14.86. Option B

18. AM = (46+79+26+85+39+65+99+29+56+72)/10 = 59.6
MD = (13.6+19.4+33.6+25.4+20.6+5.4+39.4+30.6+3.6+12.4)/10 = 204/10 = 20.4
Coefficient of MD = 20.4/59.6 * 100 = 34.23% Option C

19. COV = 6.25/25 * 100 = 25. Option B

20. $\sum X$ = 12*20 = 240
Correct $\sum X$ = 240 – 15 – 27 + 11 + 21 = 230
$\sum X^2$ = 20 [36 + 144] = 3600
Correct $\sum X^2$ = 3600 – 15^2 – 27^2 + 11^2 + 21^2 = 3600 – 225 – 729 + 121 + 441 = 3208
Correct SD^2 = 3208/20 – (230/20)^2 = 160.4 – 132.25 = 28.15
SD = $\sqrt{28.15}$ = 5.3. Option C

21. 5/100 = 2/Mean. Mean = 200/5 = 40. Option B

22. Mean = 4.4 * 5 = 22
Sum of remaining two numbers = 22 – 1 – 2 – 6 = 13. Only option fits in is Option B

23. SD2 = 200/10 – (20/10)2 = 20 – 4 = 16. SD = 4. Option B

24. SD(y) = 3.SD(x) = 3*4 = 12. Option B

25. After arrangement: 10, 12, 14, 17, 19, 21, 27, 28, 30, 32, 34
Q3 = ¾(11+1) = 9th item = 30
Q1 = ¼(11+1) = 3rd item = 14
QD = (Q3 – Q1)/2 = (30 – 14)/2 = 16/2 = 8. Option A

26. SD = ½ |16 − 10| = 6/2 = 3. Option C

27. Coefficient of Range = (H − L)/(H + L) = (60 − 0)/(60 + 0) = 1. Option A

28. 0.37 = (46.25 − Q1)/(46.25 + Q1)
    17.1125 + 0.37Q1 = 46.25 − Q1
    1.37Q1 = 29.1375
    Q1 = 29.1375/1.37 = 21.26. Option B

29. Sum of squares = 200[9 + 482) = 462600
    Option A

30. Combined Mean = (50*54.4 + 100*50.3)/150 = 51.7
    D1 = 54.4 − 51.7 = 2.7, D2 = 50.3 − 51.7 = - 1.4
    Combined SD2 = [50(82 + 2.72) + 100(72 + 1.42)]/150 = 57.737
    Combined SD = (57.737)1/2 = 7.59
    Option C