

In the name of SMART-WORK
Don't invent intelligent ways
to escape HARD-WORK

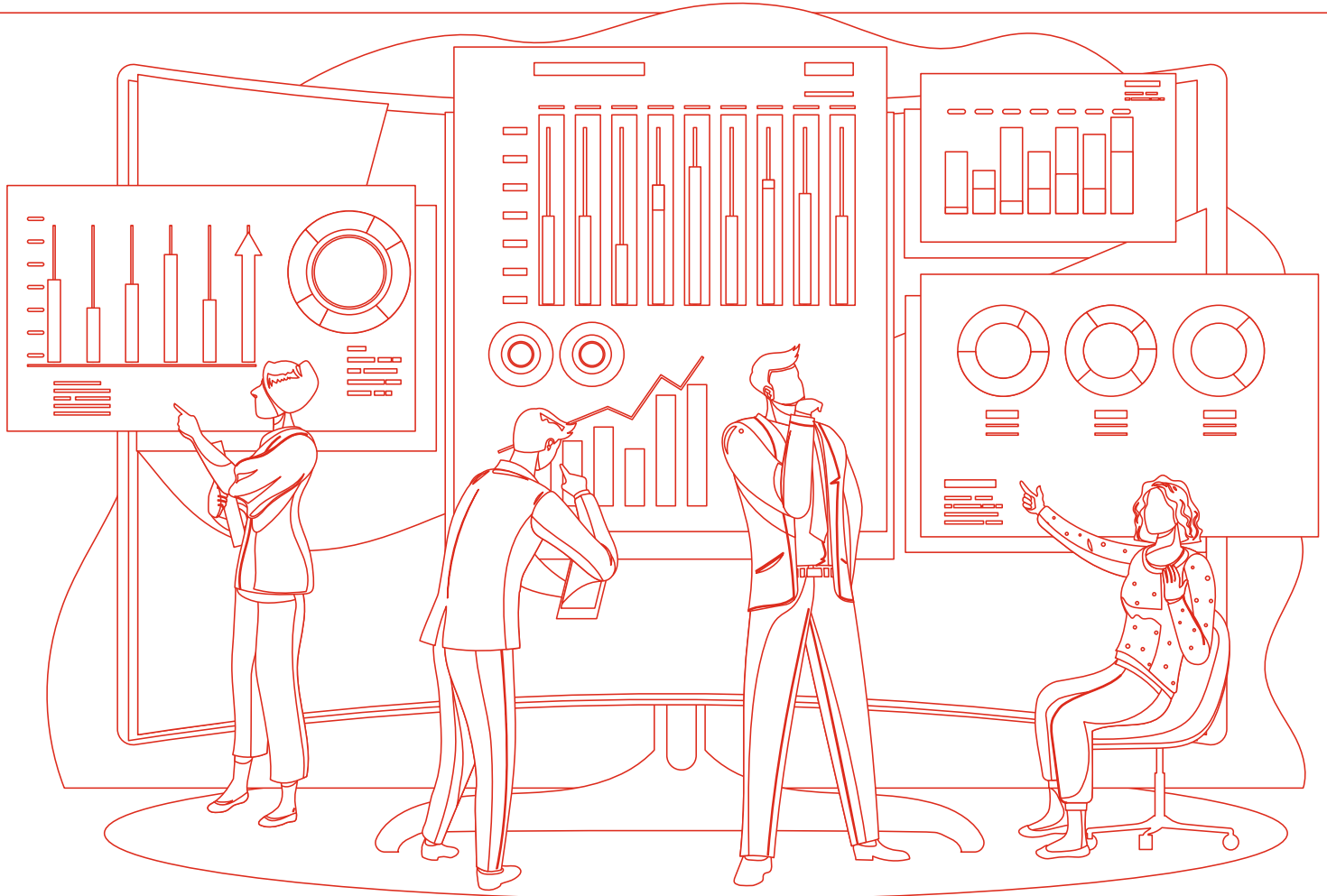
- CA VINOD REDDY -



Before you Work SMART You must Work HARD

- CA VINOD REDDY -





Statistical Description of Data

CA Vinod Reddy

1 The word statistics is derived from :

Latin word Status

Italian word Statista

German word Statistik

French word Statistique

2 We may define statistics in singular and plural sense

3 Statistics is useful in -

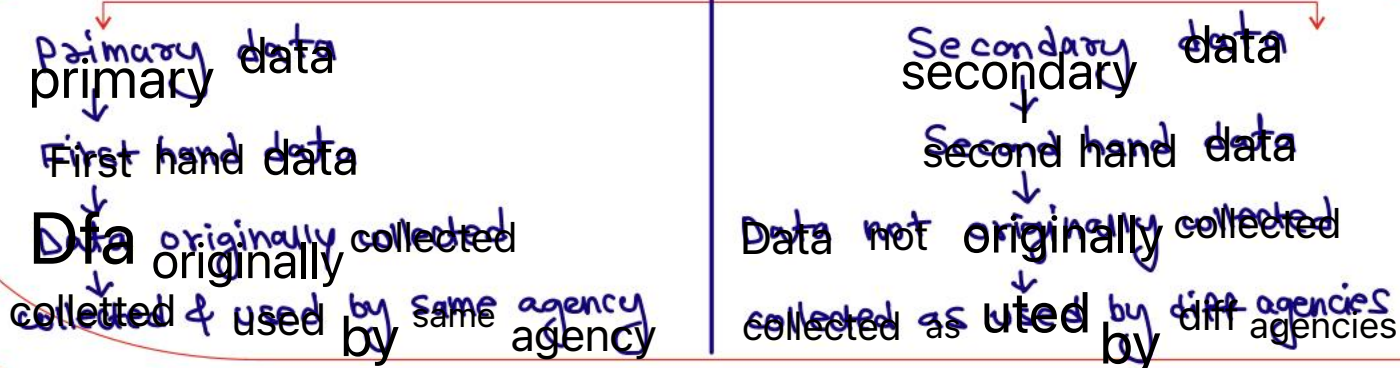
- ① Business Economics
- ② Business management
- ③ commerce & related fields
- ④ medical, engineering fields
- ⑤ Investment managements
- ⑥ Insurance, portfolio mgmt
- ⑦ politics, sports etc.

4 5 Steps in Statistics -

- ① Collection of data
- ② Scrutiny of data
- ③ Classification of data
- ④ Presentation of data
- ⑤ Analysis of data

(CSECPA)

5 Collection of Data



6 Following methods can be used for collection of primary data

1. Questionnaire Method
2. Mailed questionnaire Method
3. Interview Method
4. Observation Method

7 Sources of Secondary Method

1. International sources WHO, IMF, World Bank, etc.
2. Govt. Sources
3. Private Sources
4. Unpublished Data etc

8 Checking the data for its Accuracy and consistency is known as scrutiny of data

9 Methods of Classification of Data

1. Geographical classification
2. Chronological classification (Temporal classification)
3. Qualitative classification
4. Quantitative classification
5. _____

10 Methods of Presentation of Data

1. Textual presentation : Most lengthy method
2. Tabular presentation : Best method
3. Diagrammatic presentation : most attractive method

11 Table No. 678 Course wise No. of students at PERCEPT (Year 2022) → Title

table number captions (Year 2022) → Head - Note

Course	Boys	Girls	Total
CA Foundation	2250	200	450
CA Inter	1000	1200	2200
CA Final	900	200	1100
Total	3,150	1,600	3,750

Foot Note → source : www.perceptforca.com

12 The best method of data presentation is

Tabular presentation

13 The most attractive method of data presentation is

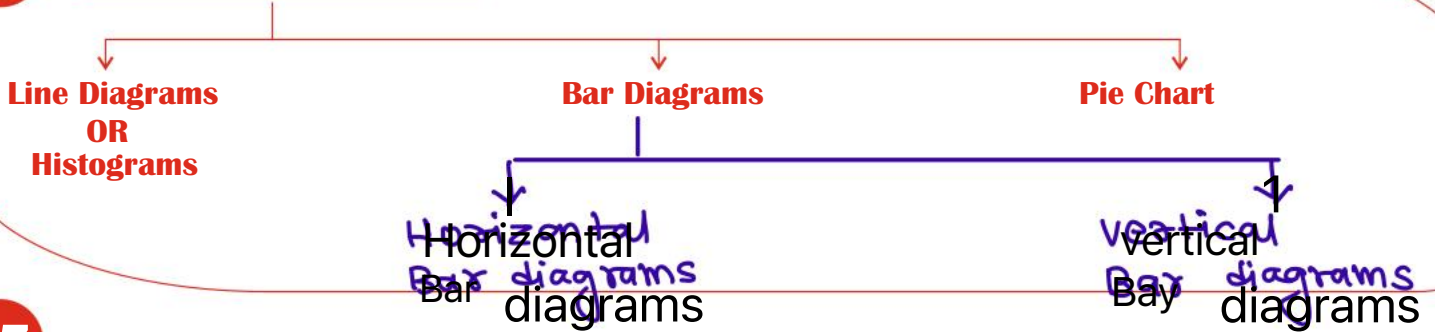
Diagrammatic presentation → Graphs, pie charts, Histograms, Bar diagrams, ogives, comma free curves, pictorial presentation.

14 Stubs are : Tree Diagrams

Left part of the table describing the rows.

15 Captions are :
 upper part of table describing columns, sub columns.

16 Diagrammatic Presentation



17 Simple data on marks of 20 students :
 6, 3, 8, 11, 19, 23, 24, 18, 11, 13, 16, 15, 19, 11, 20, 16, 8, 9, 2, 3, 5, 4, 9, 2, 13

Grouped & Classified data

C.I.	Frequency (no. of times an observation occurs)
0-5	= 5
5-10	= 6
10-15	= 5
15-20	= 6
20-25	= 3

$\sum f = N = 25$

18 Tally marks are used represent frequency
 8, 9, 8, 10, 8, 10, 10, 9, 10, 10, 9, 8, 8, 8, 10, 10, 10, 9, 8

se	f
8	7
9	4
10	8
	19

simple data
 Grouped data
 $\therefore \sum f = N = 19$

19 $LCB = LCL - (D/2)$

$UCB = UCL + (D/2)$

Relative Frequency = $(\text{Freq. of the class OR these } f / \text{Total Frequency})$

Percentage Frequency = $(\text{Freq. of the class OR these } f / \text{Total Frequency}) \times 100$

Class Width = $(UCB - LCB)$

Class-mark = $(\frac{LCL + UCL}{2}) = (\frac{LCB + UCB}{2})$

Frequency Density = $(\text{Freq. of the class} / \text{class width})$

Less than type of cumulative frequency = $[\text{Freq. of the class} + \text{sum of freq. of all preceding classes}]$

Greater than type of cumulative frequency = $[\text{If } f_s + \text{sum of freq. of all succeeding classes}]$

20 exclusive classification : preferred for continuous variable

C.I.	Freq.	LCL	UCL	LCB	UCB	Relative Freq.	% Freq.	Freq. Density	Class Mark	Class Width	less than type c.f.	greater than type c.f.
10-20	5	10	20	10	20	0.10	10%	0.50	15	10	5	50
20-60	8	20	60	20	60	0.16	16%	0.20	40	40	13	95
60-80	7	60	80	60	80	0.14	14%	0.35	70	20	20	37
80-100	20	80	100	80	100	0.40	40%	1.00	90	20	40	30
100-120	3	100	120	100	120	0.06	6%	0.15	110	20	43	10
120-140	7	120	140	120	140	0.14	14%	0.35	130	20	50	7

Inclusive classification : preferred for discrete variable

21

C.I.	Freq	LCL	UCL	LCB	UCB	Relative Freq.	% Freq.	Class Mark	Class Width	less than type c.f.	greater than type c.f.
10-18	2	10	18	9	19	0.04	4%	14	10	2	50
20-38	8	20	38	19	39	0.16	16%	29	20	10	48
40-98	10	40	98	39	99	0.20	20%	69	60	20	40
100-168	13	100	168	99	169	0.26	26%	134	70	33	30
170-218	8	170	218	169	219	0.16	16%	194	50	41	17
220-318	5	220	318	219	319	0.10	10%	269	100	46	9
320-398	4	320	398	319	399	0.08	8%	359	80	50	4

22 Graphical Presentation

Area Diagrams OR Histograms

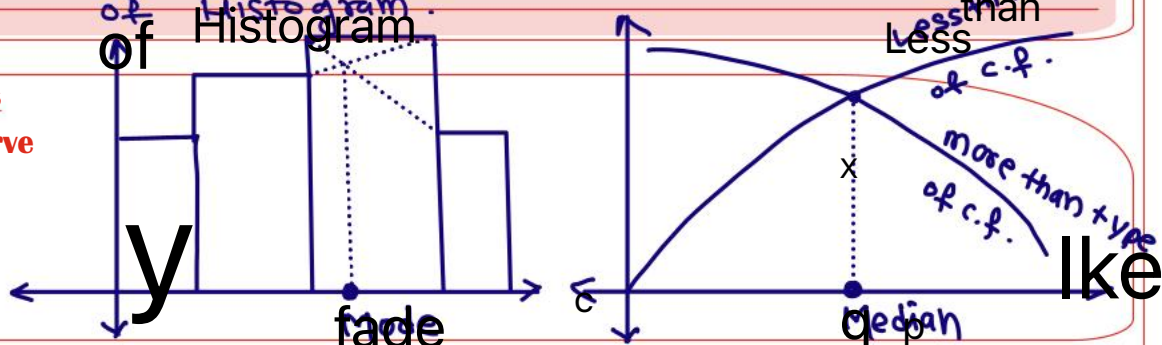
Frequency Polygon

Cumulative frequency curves OR Ogives

23 Median can be graphically obtained with the help of cumulative frequency curves or ogives.

Mode can be graphically obtained with the help of Histogram.

- 24** Frequency Curves
1. Bell shaped curve
 2. U-shaped curve
 3. J-shaped curve
 4. Mixed curve



25 Data

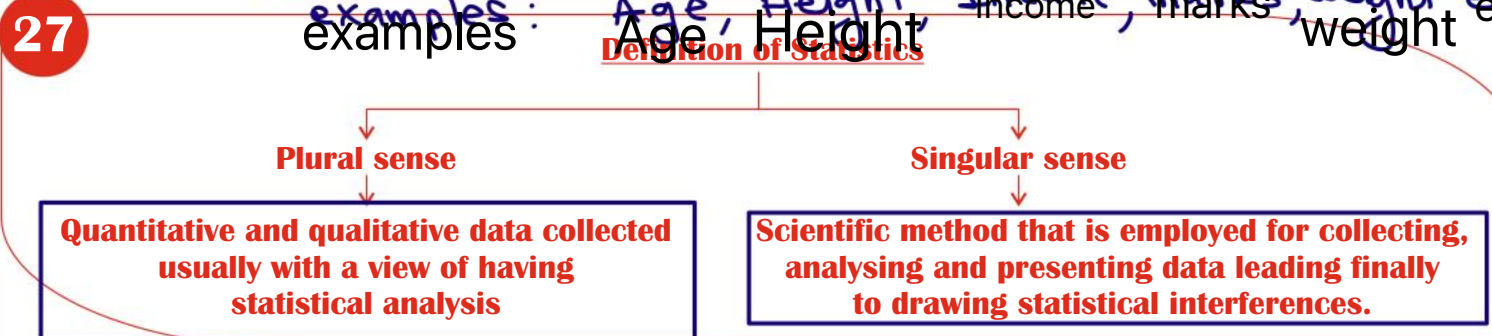
A Qualitative characteristic is known as attribute

A Quantitative characteristic is known as variable

- ① Discrete variable
- ② continuous variable

26 **Discrete Variable :** Within a given interval, if a variable can assume finite/limited values, It is said to be a Discrete variable. examples: No. of students, no. of errors in a book

Continuous Variable : Within a given interval if a variable can assume infinite/unlimited values, It is said to be a continuous variable. examples: Age, Height, Income, marks, weight etc



- 28** **Limitation of Statistics**
- a. Deals with aggregate, an individual has no statistical significance.
 - b. Mostly concerned with quantitative data
 - c. Based on assumptions, so projections are likely to be inaccurate
 - d. Based on random sampling.

29 **Methods of Collection of primary data**

Interview Method	Mailed Questionnaire Method	Observation Method	Questionnaire filled and sent by enumerators
a. Personal interview			
b. Indirect interview			
c. Telephonic interview			

- 30**
- a. In personal interview investigator meets to the respondent directly and collects the information.
 - b. If there are some practical problems in reaching the respondents directly then we may go for indirect interview when investigator collects the information from the persons associated with the problem.
 - c. Telephonic interview is quick and non expensive method to collect primary data.
- First 2 methods are inapplicable when there is large data. The amount of non-response is maximum for third method of data collection.

My Notes

$$L_{CB} = L_{CL} - (D/E)$$

$$U_{CB} = U_{CL} + (D/E)$$

$$\text{Freq. density} = \left(\frac{\text{Freq. of the class}}{\text{class width}} \right)$$

	f	Relative Freq.	% Freq.
Boys	70	0.35	35%
Girls	130	0.65	65%
		1.00	100%

31

Scrutiny of Data :

Since statistical analysis are made only on the basis of data, it is necessary to check whether the data under consideration are accurate and consistent.

No hard and fast rules can be applied for scrutiny of data. One must apply his intelligence, patience and experience while scrutinising the given information.

32

Textual Presentation :

This method comprises presenting data with the help of paragraphs.

Advantage of this method lies in its simplicity, a layman can also present data under this method.

Textual presentation, however not preferred as it is Dull, Monotonous, Lengthy.

33

Tabular Presentation :

It may be defined as systematic presentation of data with the help of a statistical table having no. of rows, columns and complete ref. no., title, description of rows and columns, foot notes, if any.

a. Caption is the upper part of the table describing column and sub-columns.

b. Stubs are left part of table providing description of rows.

c. Body is the main part of the table that contains numerical figures.

d. Addi. info, Disclaimer, source of data in mentioned in : Foot Note

e. units of measurement can be written in : Box-Head.

34

Diagrammatic Presentation of Data

a. Another alternative and attractive method is with the help of charts, graphs, pictures, etc.

b. Any hidden trend can be understood with the help of this method.

c. However, as compared to tabulation, this method is less accurate. So if priority is accuracy of data, we have to recommend tabulation.

• The Best method of data presentation is : Tabular presentation

35

We are going to consider the following types of diagrams

a. Line diagram / histogram

b. Bar diagram

c. Pie chart / pie diagram / circle diagram.

My Notes

Stubs are : Left part of table describing rows

captions are : upper part of table describing columns & sub-columns.

36

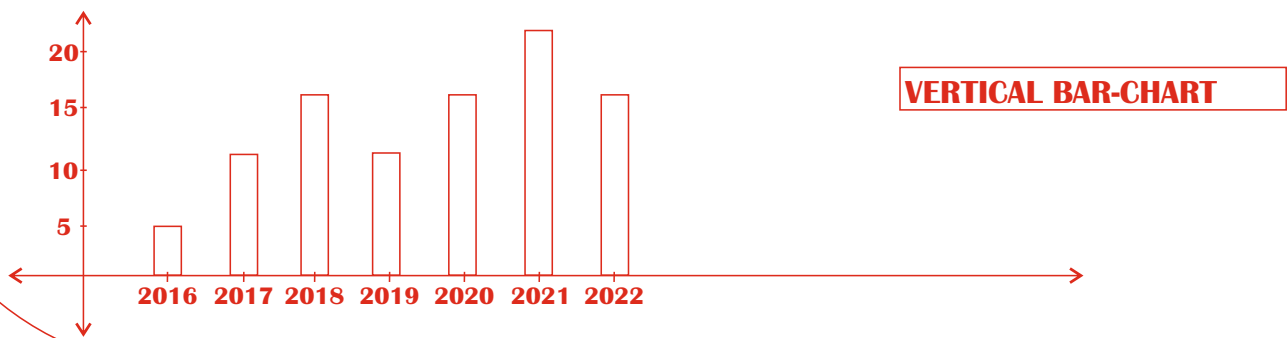
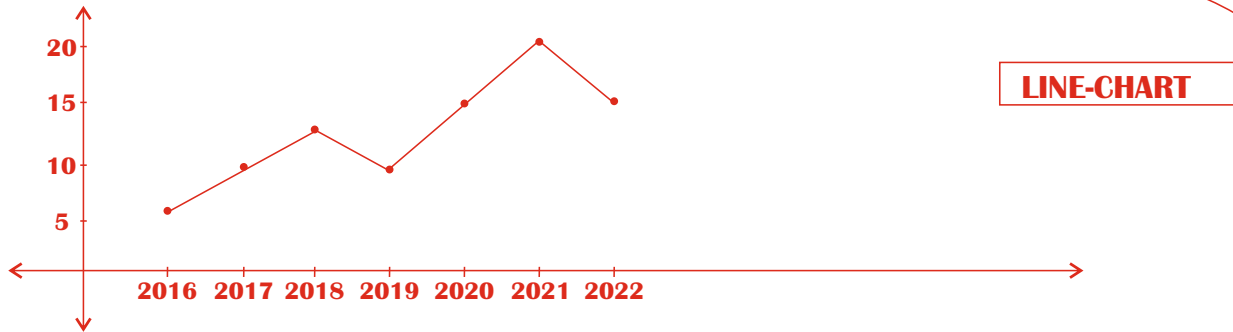
- Line diagram that uses logs is known as Ratio-chart.
- Multiple Line chart is used for representing 2 or more related time series data expressed in same unit.
- Multiple Axis chart in somewhat similar situations if variables are expressed in different units.

37

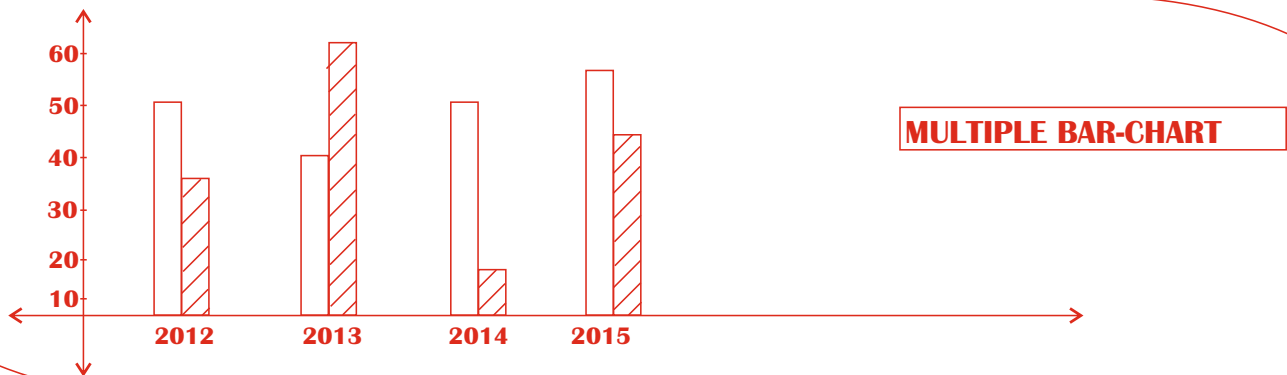
- Horizontal bar diagram issued for qualitative data.

- Vertical bar diagram is associated with quatitative data OR time series data

38



39



My Notes

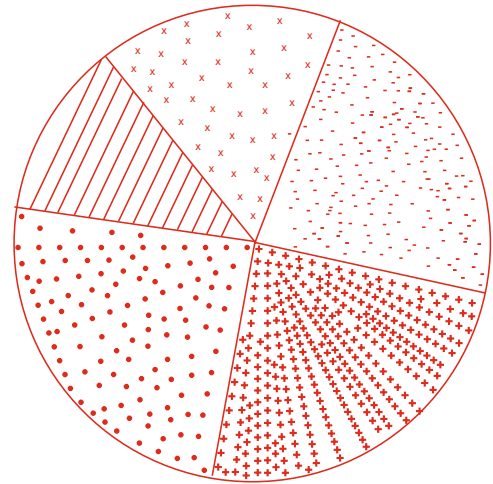
40 Draw the appropriate diagram for presentation the of following data :

Source	Revenue in Millions (₹)
Customs	80
Excise	190
Income-Tax	160
Corporate Tax	75
Misc	35
Total	540

Angle in a pie-chart

$$= \left(\frac{\text{value of a component}}{\text{Total Value}} \right) \times 360^\circ$$

Source	Angle in Pie chart
Customs	$(80/540) \times 360 = 53^\circ$ (approx.)
Excise	$(190/540) \times 360 = 127^\circ$
Income-Tax	$(160/540) \times 360 = 107^\circ$
Corporate Tax	$(75/540) \times 360 = 50^\circ$
Misc	$(35/540) \times 360 = 123^\circ$



Pie-chart OR Pie-Diagram

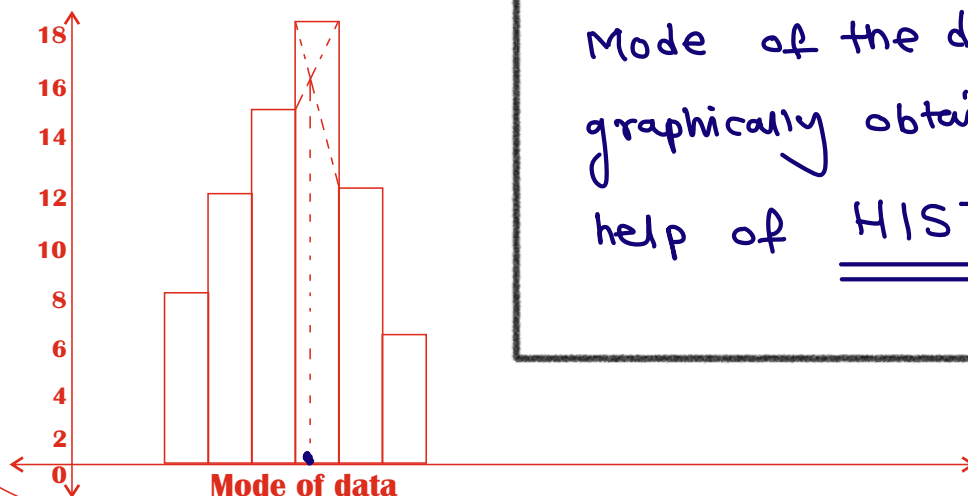
41 Graphical Presentation of Frequency Distribution

Area Diagrams
OR
Histograms

Frequency Polygon

Cumulative frequency curves
OR
Ogives

42 Histogram

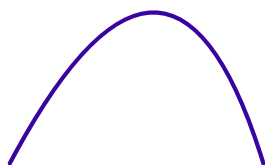


Mode of the data can be graphically obtained with the help of HISTOGRAM

45 There are 4 types of frequency curves

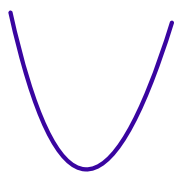
Bell shaped curve

Distribution of height, weight income generally belong this category. Freq. distribution starts with low becomes maximum at centre then gradually reaches to lowest value to other extremity.



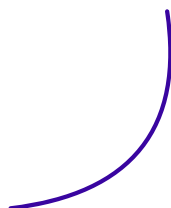
U shaped curve

Frequency distribution is minimum near the central part and freq. slowly and steadily reaches maximum at after two extremities.



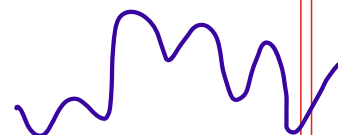
J shaped curve

It starts with the minimum frequency and then gradually reaches to maximum frequency at other extremity.



Mixed curve

We may have combination of these frequency curves. No specific shape for mixed curve.



46 The primary data is collected by

a. Interview Method

b. Observation Method

c. Questionnaire Method

d. All of these

47 The quickest method to collect primary data is :

a. Personal Interview

b. Indirect Interview

c. Telephonic interview

d. Observation Method

My Notes

scrutiny of data

Checking the data for its accuracy & consistency & Authenticity is known as Scrutiny of data

48 In case of Rail accident, the appropriate method of data collection is by :

a. Personal Interview

b. Direct Interview

c. Indirect Interview

d. All of these

49 Which method of data collection covers widest area

a. Telephonic interview

b. Mailed Questionnaire Method

c. Direct Interview Method

d. All of these

50 The amount of non-responses are maximum in case of

a. Mailed Questionnaire Method

b. Interview Method

c. Observation Method

d. All of these

51 The accuracy and consistency of data can be verified by -

a. Internal checking

b. External checking

c. Scrutiny

d. None of these

52 The unit of measurement in tabulation is shown in the

a. Box Head

b. Body

c. Caption

d. Stub

53 In tabulation, source of the data if any is shown in the

a. Foot-Note

b. Body

c. Caption

d. Stub

54 Hidden trend, if any, in a data can be noticed by

a. Textual presentation

b. Tabulation

c. Diagrammatic Presentation

d. None of these

55 The most accurate (Best) method of data presentation is :

a. Diagrammatic Presentation

b. Tabulation

c. Textual presentation

d. None of these

My Notes

56 The chart used logarithms of a variable is known as :

- a. Line chart
- b. Ratio chart
- c. Multiple line chart
- d. Pie chart

57 Pie diagram is used for?

- a. Comparing diff. components and their relation to total
- b. Representing qualitative data in a circle
- c. Representing quantitative data in a circle
- d. b or c

58 A frequency distribution

- a. Arranges observations in increasing order
- b. Arranges observations in number of groups
- c. is for time pass
- d. All of these

59 Frequency distribution of a continuous variable is known as

- a. Grouped frequency distribution
- b. Simple frequency distribution
- c. a or b
- d. a and b

60 The distribution of shares is an example of frequency distribution of :

- a. A discrete variable
- b. A continuous variable
- c. An attribute
- d. None of these

61 The distribution of profits of a blue chip company relates to :

- a. A discrete variable
- b. A continuous variable
- c. An attribute
- d. None of these

62 Mutually exclusive classification

- a. Excludes both the class limits
- b. Excludes UCL but includes LCL
- c. Includes UCL and excludes LCL
- d. None of these

10-20
20-30 includes LCL & excludes UCL
 (20) (30)
 30-40

My Notes

- For exclusive classification,
 $L C B = L C L$
 $U C B = U C L$

63 Out of 1000 workers, 25% were industrial workers and rest were agricultural workers. 300 persons enjoyed world cup matches on T.V, 30% of people who had not watched world cup matches were industrial workers. What is agri. no. of workers who had enjoyed world cup matches on T.V.?

- a. 260 b. 240 c. 230 d. 250

workers	Industrial	Agri.	Total
watched world cup on TV	40	260	300
Not watched	210	490	700
Total	250	750	1,000

Textual to Tabular presentation

64 The number of accident for 7 days in a locality are given below :

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

What is no. of cases when 3 or less accidents occur?

- a. 56 b. 6 c. 68 ~~d. 87~~

$$15 + 19 + 22 + 31 = 87 \text{ cases}$$

65 The follow data relates to income :

Income	500 - 999	1000 - 1499	1500 - 1999	2000 - 2499
No. of persons	15	28	36	7

What is % of persons earning more than ₹ 1500?

- a. 43% ~~b. 50%~~ c. 40% d. None of these

$$(36 + 7) = 43 \quad \left(\frac{43}{86} \times 100 \right) = 50\%$$

66 The following data relate to the marks of group of students :

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

How many students have marks more than 30?

- a. 65 b. 184 ~~c. 35~~ d. None of these

My Notes

c.I. (marks)	No. of students (f)
0-10	15
10-20	23
20-30	27
30-40	19
40-50	16

$$19 + 16 = 35$$

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

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

- a. 56 ~~b. 23~~ c. 15 d. 8

C.I.	frequency
200 - 250	56 - 38 = 18
250 - 300	38 - 15 = 23
300 - 350	15 - 0 = 15

68 Cost of sugar in a month under the heads material, labour, expenses, overheads are ₹ 12,20,35,23 respectively. What is diff between central angles for the largest and smallest components of cost of sugar?

- a. 72° b. 48° c. 56° ~~d. 92°~~

$$\begin{aligned} \text{Largest Angle} &= \frac{35}{90} \times 360 = 140^\circ & \text{diff} \\ \text{smallest Angle} &= \frac{12}{90} \times 360 = 48^\circ & = 140 - 48 \\ & & = 92^\circ \end{aligned}$$

69 The distribution of profits of a company generally follows :

- a. J shaped freq. curve b. U-shaped frequency curve
~~c. Bell shaped freq. curve~~ d. None of these

70 The distribution most commonly used is :

- a. Mixed b. U-shaped ~~c. Bell shaped~~ d. None of these

71 Graph is a _____

- ~~a. Line diagram~~ b. Bar-diagram c. Pie-diagram d. Pictogram

72 (Class frequency / class width) is defined as _____

- ~~a. Frequency density~~ b. Frequency distribution c. Both d. None

My Notes

73 Tally Marks determines

- a. Class width
- b. Class boundary
- c. Class limit
- d. Class Frequency

74 An area diagram is

- a. Histogram
- b. Frequency Polygon
- c. Ogives
- d. None

75 Ogive is a _____

- a. Line diagram
- b. Bar diagram
- c. Both
- d. None

76 Unequal width of classes in a frequency distribution do not cause any difficulty in construction of _____

- a. Ogive
- b. Frequency Polygon
- c. Histogram
- d. None of these

77 Graphical presentation of cumulative frequency distribution is called as _____

- a. Histogram
- b. Ogive
- c. Both
- d. None of these

78 The most common form of diagrammatic presentation of a grouped frequency distribution is

- a. Ogive
- b. Histogram
- c. Frequency Polygon
- d. None of these

79 Vertical Bar diagram may appear somewhat alike -

- a. Histogram
- b. Frequency Polygon
- c. Ogive
- d. None of these

80 Number of types of cumulative frequency is :

- a. One
- b. Two
- c. Three
- d. Four

My Notes

81 A representative value of a class interval for the calculation of Mean, SD, MD, etc. is

- a. Class interval
- b. Class limit
- c. Class mark
- d. None

82 In all statistical calculations & diagrams involving end points of classes _____ are used.

- a. Class Boundaries
- b. Class Values
- c. both
- d. None

83 Upper boundary of a class coincide with Lower boundary of next class.

- a. True
- b. False
- c. Both
- d. None

84 The lower extreme point of a class is called as _____

- a. Lower Class Limit
- b. Lower Class Boundary
- c. Both
- d. None

Below 10
10-30
30-80
80-above

} Data with open class intervals

85 When one end of the class is not specified, the class is called as _____

- a. Open end class
- b. Close end class
- c. Both
- d. None of these

86 When all classes have equal width, the heights of rectangles in histogram will be numerically equal to the _____ .

- a. Class Frequencies
- b. Class Boundaries
- c. Both
- d. None of these

87 To find 'Mode of data' graphically we use _____ .

- a. Ogives
- b. Frequency Polygon
- c. Histogram
- d. None of these

88 In representing simple frequency distributions of a discrete variable _____ is useful.

- a. Ogives
- b. Histogram
- c. Frequency Polygon
- d. None of these

My Notes

89 Diagrammatic presentation of cumulative frequency distribution is _____ .

a. Frequency Polygon

b. Ogives

c. Histogram

d. None of these

90

Class	0-10	10-20	20-30	30-40	40-50
Frequency	5	8	15	6	4

For the class 20-30 cumulative frequency is :

a. 20

b. 13

c. 15

d. 28

91 Breadth of rectangle is equal to length of class interval in _____ .

a. Ogives

b. Histogram

c. Line diagram

d. None

92 In Histogram classes are taken _____

a. Overlapping

b. None Overlapping

c. Both

d. None

93 There are _____ methods of classification of data.

a. 4

b. 3

c. 2

d. 1

94 There are _____ methods of presentation of data.

a. 4

b. 3

c. 2

d. 1

95 For the overlapping classes 0-10, 10-20, 20-30, etc. the class mark of 0 - 10 is

a. 5

b. 4.50

c. 4

d. 10

96 For the classes 0-9, 10-19, 20-29, 30-39, the class mark of 10-19 is _____

a. 14.50

b. 15

c. 20

d. 16

97 Mutually inclusive classification is meant for _____

a. Discrete variable

b. Continuous variable

c. Both

d. None

My Notes

98 Mutually exclusive classification is meant for _____

- a. Discrete variable ~~b. Continuous variable~~ c. Both d. None

99 LCB is _____

- a. Latur Crime Branch
 b. Lower Class Branch
~~c. Lower Class Boundary~~
 d. a or c

100 Relative Frequency of a particular class

- ~~a. Lies between 0 and 1~~
 b. Lies between -1 and 1
 c. Lies between -1 and zero
 d. None of these

101

Characteristic	Discrete / Continuous Variable / Attribute
a. Income	a continuous variable
b. Profit	a continuous variable
c. Blue-colour	an attribute
d. Honesty	an attribute
e. Nationality	an attribute
f. No. of shares	a Discrete variable
g. Age	a continuous variable
h. No. of members	a Discrete variable
i. Drinking habit	an attribute
j. Beauty	an attribute
k. Children in a family	a Discrete variable
l. Love	an attribute
m. Batch size	a Discrete variable

102

Class - Interval	Frequency
0 - 10	5
10 - 20	8
20 - 40	9
40 - 60	10

Mutually Exclusive Classification

Class - Interval	Frequency
0 - 9	25
10 - 29	28
30 - 89	35
90 - 189	40

Mutually Inclusive Classification

preferred for: ↓ continuous variable

↓ discrete variable

My Notes