Unit – 1 Descriptive Statistics

Course Content:

- Grouped Data and Ungrouped Data
- Mean
- Median
- Mode
- Percentiles
- Quartiles
- Five Number Summary
- Range, Inter Quartile Range
- Mean Absolute Deviation
- Population Variance, Population Standard Deviation
- Sample Variance, Sample Standard Deviation
- Frequency Distribution
- Percentage Frequency, Relative Frequency, Cumulative Frequency
- Data Graphs: Bar Graph, Histogram, Frequency Polygon, Ogive
- Stem and Leaf Display

Introduction:

Statistics can be descriptive measures computed from a sample and used to make determinations about a population.

Application of Statistics:

Statistics is useful in the areas of Marketing, Management, Finance, Economics, Accounting, Management Information Systems, and Machine Learning and in the field of A.I.

1, 0, 2, 3, 9,0.3 ,8 ,9 ,1, 1.3 ,1.4 ,5 , 6 – Ungrouped data

0- 2, 2- 4, 4 -6, 6 -8, 8- 10 – Grouped data

1) Mean:

(i) Sample Mean:

Notation : $\bar{x}$ (Read as x – bar)

Formula: $\bar{x} = \frac{\sum x}{n}$

Where $\sum x$ = Sum of total observations

n = Total number of observations (Sample Size)

(ii) Population Mean:

Notation: μ

Formula: $\mu = \frac{\sum x}{N}$

Where $\sum x$ = Sum of total observations

N = Total number of observations (Population Size)

Example 1: Find Mean of the data 1, 3, 5, 7, 9

Solution : Here n = 5

Therefore Mean $\bar{x} = \frac{\sum x}{n} = \frac{1+3+5+7+9}{5} = \frac{25}{5} = 5$

Example 2: Find Mean of the data 2, 7, 10, 14, 9, 3, 8

Here n = 7

Therefore Mean $\bar{x} = \frac{\sum x}{n} = \frac{2+7+10+14+9+3+8}{7} = \frac{53}{7} = 7.571$

2) Median :

Notation : M

To find Median first arrange data in ascending order.

Median M $= \left(\frac{n+1}{2}\right)^{th} observation$

Example 1 : Find Median of the data 6, 7, 2, 4, 1, 5

Step 1: Arrange data in ascending order

Ascending order is 1,2, 4, 5,6, 7

Here n = 6

Step 2 : M $= \left(\frac{n+1}{2}\right)^{th} observation$

M $= \left(\frac{6+1}{2}\right)^{th} observation$

M $= \left(\frac{7}{2}\right)^{th} observation$

M = 3.5th observation

$$M = \frac{3^{rd}\,observation + 4^{th}\,observation}{2}$$

$$M = \frac{4+5}{2}$$

M = 4.5

Example 2: Find Median of the data 9, 1, 2, 7, 11

Step 1: Arrange data in ascending order

Ascending order is 1, 2, 7, 9, 11

Here n = 5

Step 2 : $M = \left(\frac{n+1}{2}\right)^{th} observation$

$M = \left(\frac{5+1}{2}\right)^{th} observation$

$M = \left(\frac{6}{2}\right)^{th} observation$

M = 3rd observation

M = 7

Example 3: Find Median of the data 1, 2, 8, 9, 10, 3, 7

Ascending Order : 1, 2, 3, 7, 8, 9, 10

n = 7

Step 2 : $M = \left(\frac{n+1}{2}\right)^{th} observation$

$M = \left(\frac{7+1}{2}\right)^{th} observation$

M = M = $\left(\frac{8}{2}\right)^{th} observation$ = 4th observation = 7

Example 4: Find Median of the data 11, 9, 7, 8, 4, 12, 2, 5

Ascending Order : 2, 4, 5, 7 ,8, 9, 11, 12

n= 8

M = (4th +5th)obs / 2 = 7.5

3) Mode:

Notation : Z

If observations are repeated in a data then most frequent observation is known as Mode .

If observations are not repeated then Mode is calculated as per below

Z = 3M – 2$\bar{x}$

Where M = Median of the data

$\bar{x}$ = Mean of the data

Example 1 : Find Mode of the Data 2, 1, 3, 4, 5, 2, 4, 0, 9, 4, 6, 4, 8, 7, 12, 4, 3

Solution: Here 4 is the most frequent observation of the data.

Mode Z = 4

Example: Find Mode of the data 3, 7, 2, 12, 9, 11, 10

Solution: Here no observation is repeated. So mode is calculated as per below

$Z = 3M - 2\bar{x}$

Here n = 7

Now, $\bar{x} = \frac{\sum x}{n} = \frac{3+7+2+12+9+11+10}{7} = \frac{54}{7} = 7.7142$

Now arrange data in ascending order 2, 3, 7, 9, 10, 11, 12

Median M $= \left(\frac{n+1}{2}\right)^{th} observation$

M $= \left(\frac{7+1}{2}\right)^{th} observation$

M $= \left(\frac{8}{2}\right)^{th} observation$

M $= 4^{th}$ observation

M = 9

Now, Z = 3M – 2$\bar{x}$ = 3(9) – 2(7.7142) = 27 – 15.4248

Therefore **Z = 11.5752**

Example 1: Calculate Mean, Median and Mode of the following data:

Company	Tata	Maruti	Ford	Fiat	Honda	Nissan	Fiat	Volkswagen	BMW	Skoda
Num.of Cars (Thousand)	9.37	8.90	6.19	5.96	3.96	3.83	3.68	3.43	2.68	2.62

Solution: **(i) Mean $\overline{x} = \frac{\sum x}{n}$**

Here n = 10

$\overline{x} = \frac{\sum x}{n} = \frac{9.37+8.90+6.19+5.96+3.96+3.83+3.68+3.43+2.68+2.62}{10} = \frac{50.62}{10} = 5.062$

(ii) Median:

Arrange data in ascending order:

2.62, 2.68, 3.43, 3.68, 3.83, 3.96, 5.96, 6.19, 8.90, 9.37

Median M $= \left(\frac{n+1}{2}\right)^{th} observation$

Median M $= \left(\frac{10+1}{2}\right)^{th} observation = \left(\frac{11}{2}\right)^{th} observation = 5.5^{th}$ observation

$M = \frac{5^{th} observation + 6^{th} observation}{2} = \frac{3.83+3.96}{2} = 3.895$

(iii) Mode Z $= 3M - 2\overline{x}$

∴Z = 3(3.895) – 2(5.062) = 1.561

Exercise 1 : Find Mean, Median and Mode of the following data:

17.3   44.5   31.6   40.0,   52.8   38.8   30.1   78.5

Mean :

n = 8

$$\bar{x} = \frac{\sum x}{n} = 41.7$$

Median

Ascending Order : 17.3, 30.1, 31.6, 38.8, 40.0, 44.5, 52.8, 78.5

$$M = \left(\frac{n+1}{2}\right)^{th} observation = \left(\frac{8+1}{2}\right)^{th} observation = 4.5^{th} \text{ obs.} = (38.8 + 40.0)/2 = 39.4$$

$Z = 3M - 2\bar{x} = 3(39.4) - 2(41.7) = 118.2 - 83.4 = 34.8$

Exercise 2: Find Mean, Median and Mode for the given data:

15, 11, 14, 3, 21, 17, 22, 16, 19, 20, 4, 9

Exercise 3: Find Mean, Median and Mode for the given data:

Data	A	B	C	D	E	F	G
Value	0.732	1.274	1.2034	1.7391	2.7333	9.4322	0.1

4) Percentiles:

Percentiles are *measures of central tendency that divide a group of data into 100 parts.* There are 99 percentiles because it takes 99 dividers to separate a group of data into 100 parts.

Procedure for finding Percentiles:
 (1) Arrange data in ascending order.
 (2) Calculate Percentile Location (i) by following formula:

$$i = \frac{P}{100} \times n^{th} \text{ observation}$$

Where i = Percentile Location
P = the percentile of interest
n = total number of observations

 (3) Determine the location by either (a) or (b)

(a) If i is the whole number, the Pth percentile is the average of the value at i^{th} location and the value at $(i + 1)^{st}$ location.

(b) If i is not a whole number, the Pth percentile value is located at the $(i + 1)^{st}$ location.

It is also denoted by $\mathbf{P_i}$

P_{100} is always last observation of ascending order or it is the Largest observation of given data

Example 1:
Determine 30th percentile of the following eight numbers

14, 12, 19, 23, 5, 13, 28, 17

Solution:

Here n = 8 and P = 30

Now ascending order of the numbers is : 5, 12, 13, 14, 17, 19, 23, 28

Therefore $i = \dfrac{P}{100} \times n^{th}$ observation

$i = \dfrac{30}{100} \times 8^{th}$ observation

$i = 2.4^{th}$ observation
$i = 3^{rd}$ observation   (note: go to point (b))

$i = 13$

Example 2:
Determine P_{25} for the given data

80, 94, 105, 97, 107, 116, 112, 116, 120, 118, 119, 127

Solution :

Here n = 12 and P = 25

Now ascending order is : 80, 94, 97, 105, 107, 112, 116, 116, 118, 119, 120, 127

Therefore $i = \dfrac{P}{100} \times n^{th}$ observation

$i = \dfrac{25}{100} \times 12^{th}$ observation

i = 3rd observation

i = $\frac{3^{rd}\, observation +\, 4^{th}\, observation}{2}$ (note: go to point (a))

i = $\frac{97+105}{2} = 101$

Example 3 :
Obtain 35th Percentile and 70th Percentile of the given data

3, 5, 9, 10, 11, 12, 12, 13, 15, 15, 16, 17, 19, 19, 22, 23

Solution : (a) 35th Percentile

Here n = 16 and P = 35

Here data is already in ascending order.

Therefore i = $\frac{P}{100} \times$ nth observation

i = $\frac{35}{100} \times$ 16th observation

i= 5.6th observation

i = 6th observation
Therefore **P$_{35}$ = 12**

(b) 70th Percentile

P$_{70}$ = $\frac{70}{100} \times$ 16th observation

P$_{70}$ = 11.12th Observation

P$_{70}$ = 12th Observation

P$_{70}$ = 17

Example : 4
Find P$_{27}$, P$_{65}$, P$_{75}$ and P$_{45}$ for the following data:

22, 132, 72, 13, 9, 12, 17, 100, 102, 13, 22, 9, 5, 7, 12, 15, 9, 13, 72, 101.5, 102.6

Solution:

Ascending Order
5, 7, 9, 9, 9, 12, 12, 13, 13, 13, 15, 17, 22, 22, 72, 72, 100, 101.5 , 102, 102.6 , 132

n = 21

(i) $P_{27} = \frac{27}{100} \times 21^{\text{th}}$ observation = 5.67^{th} observation = 6th observation = 12

(ii) $P_{65} = \frac{65}{100} \times 21^{\text{th}}$ observation = 13.65^{th} observation = 14th observation = 22

(iii) $P_{75} = \frac{75}{100} \times 21^{\text{th}}$ observation = 15.75^{th} observation = 16th observation = 72

(iv) $P_{45} = \frac{45}{100} \times 21^{\text{th}}$ observation = 9.45^{th} observation = 10th observation = 13

Example: 5
Find P_{25} , P_{50}, P_{75} and P_{100} for the given data:

29, 19, 35, 27, 45, 50, 79, 81, 36, 90

Solution:
Here n = 10
Ascending Order : 19, 27, 29, 35, 36, 45, 50, 79, 81, 90

(i) $P_{25} = \frac{25}{100} \times 10^{\text{th}}$ observation = 2.5^{th} observation = 3rd observation = 29

(ii) $P_{50} = \frac{50}{100} \times 10^{\text{th}}$ observation = 5th observation = $\frac{5^{th} observation +\ 6^{th} observation}{2}$

   $= \frac{36+45}{2} = \frac{81}{2} = 40.5$

(iii) $P_{75} = \frac{75}{100} \times 10^{\text{th}}$ observation = 7.5^{th} observation = 8th observation = 79

(iv) P_{100} = 90 (Largest Observation / Last observation in the ascending order)

5) Quartiles:

Quartiles are measures of central tendency that divide a group of data into four subgroups or parts.

There are total four Quartiles denotes as Q_1 , Q_2 , Q_3 and Q_4. These Quartiles can be calculated as per below:
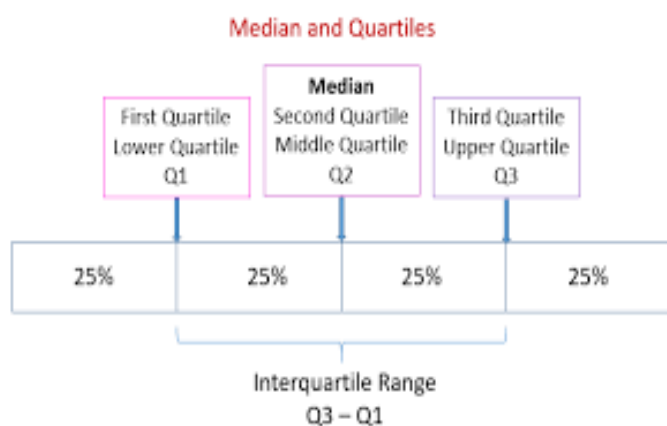
$Q_1 = P_{25}$

$Q_2 = P_{50}$

$Q_3 = P_{75}$

$Q_4 = P_{100}$

6) Inter Quartile Range (IQR):

The interquartile range is the range of values between the first and third quartile. Essentially, it is the range of the middle 50% of the data and is determined by computing the value of **Q3 - Q1**.

7) Range = L – S

where L is the Largest Value and S is the smallest value of the data

Example 1:

Determine Q_1 , Q_2 , Q_3 , Q_4 , IQR , Range and Median of the given data:

106, 109, 116, 121, 114, 125, 122, 129

Solution:

Ascending order : 106, 109, 114, 116, 121, 122, 125, 129

Here n = 8

(i) $Q_1 = P_{25} = \frac{25}{100} \times n^{th}$ observation $= \frac{25}{100} \times 8^{th}$ observation

$\qquad = 2^{nd}$ observation

$\qquad = \frac{2^{nd} observation + 3^{rd} observation}{2} = \frac{109+114}{2}$

Therefore $Q_1 = 111.5$

(ii) $Q_2 = P_{50} = \frac{50}{100} \times n^{th}$ observation $= \frac{50}{100} \times 8^{th}$ observation

$= 4^{th}$ observation

$= \frac{4^{th} observation + 5^{th} observation}{2} = \frac{116+121}{2}$

Therefore $Q_2 = 118.5$

(iii) $Q_3 = P_{75} = \frac{75}{100} \times n^{th}$ observation $= \frac{75}{100} \times 8^{th}$ observation

$= 6^{th}$ Observation

$= \frac{6^{th} observation + 7^{th} observation}{2} = \frac{122+125}{2}$

Therefore $Q_3 = 123.5$

(iv) $Q_4 = P_{100} =$ Largest Observation $= 129$

(v) IQR $= Q_3 - Q_1 = 123.5 - 111.5 = 12$

(vi) Range $= L - S = 129 - 106 = 23$

(vii) Median (M) $= \frac{(n+1)}{2}$th observation

$= \frac{8+1}{2}$th observation $= \frac{9}{2}$ th observation

$= 4.5^{th}$ observation

$= \frac{4^{th} observation + 5^{th} observation}{2} = \frac{116+121}{2}$

Therefore M $= 118.5$

Example 2 :

Determine IQR and Range of the given data:

22,195	**19,526**	**9538**	**7793**
7707	**4023**	**3916**	**3576**

3571	**3553**	**3247**	**2448**
2433	**2050**	**1137**	**699**

Solution :

Ascending Order:

699	1137	2050	2433
2448	3247	3553	3571
3576	3916	4023	7707
7793	9538	19,526	22,195

Here n = 16

(i) IQR = $Q_3 - Q_1$

Now $Q_3 = P_{75} = \frac{75}{100} \times$ nth observation $= \frac{75}{100} \times$ 16th observation = 12th Observation

$$= \frac{12^{th} observation + 13^{th} observation}{2} = \frac{7707 + 7793}{2}$$

Therefore Q_3 = 7750

And $Q_1 = P_{25} = \frac{25}{100} \times$ nth observation $= \frac{25}{100} \times$ 16th observation

$\qquad$ = 4th observation

$$= \frac{4^{th} observation + 5^{th} observation}{2} = \frac{2433 + 2448}{2}$$

Therefore Q_1 = 2440.5

Now, **IQR = $Q_3 - Q_1$ = 7750 – 2440.5 = 5309.5**

(ii) Range = L – S = 22,195 – 699 = 21, 496

Exercise : 3

Compute the 35th percentile, the 55th percentile, Q1, Q2, and Q3 for the following data.
16 28 29 13 17 20 11 34 32 27 25 30 19 18 33

Exercise: 4

Determine IQR for the following data:

213 345 609 73 167 243 444 524 199 682

Exercise: 5

Determine all four Quartiles of the given data:

80, 94, 97, 105, 107, 112, 116, 116, 118, 119, 120, 127, 128, 138, 138, 139, 142, 143, 144, 145, 150, 162, 172, 172

8) Five Number Summary:

A five-number summary is especially useful in descriptive analyses or during the preliminary investigation of a large data set. A summary consists of five values:

(i) Smallest Value (S) (ii) First Quartile (Q₁) (iii) Median (M) (iv) Third Quartile (Q₃) (v) Largest Value (L)

These values have been selected to give a summary of a data set because each value describes a specific part of a data set: the median identifies the centre of a data set; the upper and lower quartiles span the middle half of a data set; and the highest and lowest observations provide additional information about the actual dispersion of the data. This makes the five-number summary a useful measure of spread.

Example 1: Obtain five number summary of the following data:

Company	A	B	C	D	E	F	G	H
Revenue	66.11	47.41	43.81	40.19	21.82	19.86	17.27	7.88

Solution: Here n = 8

Ascending order : 7.88, 17.27, 19.86, 21.82, 40.19, 43.81, 47.41, 66.11

(i) Smallest Value (S) = 7.88

(ii) First Quartile $(Q_1) = P_{25} = \frac{25}{100} \times n^{th}$ observation $= \frac{25}{100} \times 8^{th}$ observation

$$= 2^{nd} \text{ observation}$$

$$= \frac{2^{nd}\,observation + 3^{rd}\,observation}{2} = \frac{17.27 + 19.86}{2}$$

First Quartile (Q_1) = 18.565

(iii) Median M $= \frac{(n+1)}{2}$th observation

$$= \frac{8+1}{2}\text{th observation} = \frac{9}{2} \text{ th observation}$$

$$= 4.5^{th} \text{ observation}$$

$$= \frac{4^{th}\,observation + 5^{th}\,observation}{2} = \frac{21.82 + 40.19}{2}$$

Therefore M = 31.005

(iv)Third Quartile (Q_3)= $P_{75} = \frac{75}{100} \times n^{th}$ observation $= \frac{75}{100} \times 8^{th}$ observation

$$= 6^{th} \text{ Observation}$$

$$= \frac{6^{th}\,observation + 7^{th}\,observation}{2} = \frac{43.81 + 47.41}{2}$$

Therefore Q_3 = 45.61

(v) Largest Value (L) = 66.11

Example 2: Determine five number summary of the following data

12, 72, 13, 91, 11, 34, 56, 79, 87, 65, 56, 39, 63, 82, 95, 101

Solution : Here n = 16

Now ascending order is

11, 12, 13, 34, 39, 56, 56, 63, 65, 72, 79, 82, 87, 91, 95, 101

(i) Smallest Value (S) = 11

(ii) First Quartile (Q_1) = P_{25} $= \frac{25}{100} \times n^{th}$ observation $= \frac{25}{100} \times 16^{th}$ observation

$$= 4^{th} \text{ observation}$$

$$= \frac{4^{th}\,observation + 5^{th}\,observation}{2} = \frac{34 + 39}{2}$$

First Quartile (Q_1) = 36.5

(iii) Median (M) $= \frac{(n+1)}{2}$ th observation

$$= \frac{16+1}{2} \text{th observation} = \frac{17}{2} \text{ th observation}$$

$$= 8.5^{\text{th}} \text{ observation}$$

$$= \frac{8^{th} observation + 9^{th} observation}{2} = \frac{63+65}{2}$$

Median (M) = 64

(iv) Third Quartile (Q_3)= $P_{75} = \frac{75}{100} \times n^{\text{th}}$ observation $= \frac{75}{100} \times 16^{\text{th}}$ observation

$$= 12^{\text{th}} \text{ Observation}$$

$$= \frac{12^{th} observation + 13^{th} observation}{2} = \frac{82+87}{2}$$

Therefore Q_3 = 84.5

(v) Largest Value (L) = 101

9) Mean Absolute Deviation:

The **mean absolute deviation (MAD)** is the average of the absolute values of the deviations
around the mean for a set of numbers.

Formula : MAD $= \frac{\Sigma |x - \bar{x}|}{n}$

Where $\bar{x} = \frac{\Sigma x}{n}$

n = Total number of observations

Σx = Sum of total observations

Example 1 :

Calculate Mean Absolute Deviation of the following data:

3, 7, 11, 15, 2, 12, 20 ,25 , 13, 15

Solution :

MAD $= \frac{\Sigma |x - \bar{x}|}{n}$

Here n = 10

Now $\bar{x} = \frac{\sum x}{n} = \frac{3+7+11+15+2+12+20+25+13+15}{10} = \frac{123}{10} = 12.3$

x	x - $\bar{x}$	| x - $\bar{x}$ |
3	$3 - 12.3 = -9.3$	9.3
7	$7 - 12.3 = -5.3$	5.3
11	$11 - 12.3 = -1.3$	1.3
15	$15 - 12.3 = 2.7$	2.7
2	$2 - 12.3 = -10.3$	10.3
12	$12 - 12.3 = -0.3$	0.3
20	$20 - 12.3 = 7.7$	7.7
25	$25 - 12.3 = 12.7$	12.7
13	$13 - 12.3 = 0.7$	0.7
15	$15 - 12.3 = 2.7$	2.7
		$\sum \| x - \bar{x} \| = 53$

Therefore MAD $= \frac{\sum|x - \bar{x}|}{n} = \frac{53}{10} = \mathbf{5.3}$

Example 2: Find MAD of the given data:

55, 100, 125, 140, 60

Solution :

MAD $= \frac{\sum|x - \bar{x}|}{n}$

Here n = 5

Now $\bar{x} = \frac{\sum x}{n} = \frac{55+100+125+140+60}{5} = \frac{480}{5} = 96$

x	x - $\bar{x}$	| x - $\bar{x}$ |

55	55- 96 = -41	41
100	100-96 = 4	4
125	125 – 96 = 29	29
140	140 – 96 = 44	44
60	60- 96 = -36	36
		$\sum \mid x - \bar{x} \mid = 154$

Therefore MAD $= \frac{\Sigma|x-\bar{x}|}{n} = \frac{154}{5} = 30.8$

Example 3: Find MAD and Five Number Summary of the given data

3, 5, 6, 7, 8, 8, 9, 10, 11, 12, 13, 14

Example 4 : Find MAD and IQR of the given data:

2, 4, 6, 8, 10, 12, 13, 20, 21, 25

Solution : Here n = 10

(i)Now IQR = Q$_3$ – Q$_1$ = P$_{75}$ – P$_{25}$

Now Third Quartile Q $_3$ =P$_{75}$ = $\frac{75}{100} \times$ nth observation = $\frac{75}{100} \times 10^{th}$ observation

$\qquad$ = 7.5th observation

$\qquad$ = 8th observation

Therefore Q$_3$ = 20

Now, First Quartile Q$_1$ = P$_{25}$ = $\frac{25}{100} \times$ nth observation = $\frac{25}{100} \times 10^{th}$ observation

$\qquad\qquad$ = 2.5th Observation

$\qquad\qquad$ = 3rd Observation

First Quartile (Q$_1$) = 6

Therefore IQR = Q$_3$ – Q$_1$ = 20 – 6 = 14

(ii) MAD $= \frac{\Sigma|x - \bar{x}|}{n}$

Now $\bar{x} = \frac{\Sigma x}{n} = \frac{2+4+6+8+10+12+13+20+21+25}{10} = \frac{121}{10} = 12.1$

x	x - $\bar{x}$	| x - $\bar{x}$ |
2	2 – 12.1 = -10.4	10.4
4	4 – 12.1 = -8.1	8.1
6	6 – 12.1 = -6.1	6.1
8	8 – 12.1 = -4.1	4.1
10	10 – 12.1 = -2.1	2.1
12	12 – 12.1 = -0.1	0.1
13	13 – 12.1 = 0.9	0.9
20	20 – 12.1 = 7.9	7.9
21	21 – 12.1 = 8.9	8.9
25	25 – 12.1 = 12.9	12.9
		Σ| x $-$ $\bar{x}$ | = 61.5

Now, **MAD** $= \frac{\Sigma|x - \bar{x}|}{n} = \frac{61.5}{10} = \textbf{6.15}$

**<u>10)Population Variance, Population Standard Deviation, Sample Standard
Deviation and Sample Variance</u>**

**The variance is the average of the squared deviations about the arithmetic
mean for a set
of numbers.**

(i) Population Variance :

$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$

Where μ = Population Mean

N = Population Size

(ii) Sample Variance:

$$S^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

Where $\bar{x}$ = Sample Mean

n = Sample Size

(iii) Population Standard Deviation:

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

(iv) Sample Standard Deviation:

$$S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

Example 1 : Find Population Variance and Sample Variance of the given data:

5, 9, 16, 17, 18

Solution:

Here n = 5

Now $\bar{x} = \frac{\sum x}{n} = \frac{5+9+16+17+18}{5} = \frac{65}{5} = 13$

x	$x - \bar{x}$	$(x-\bar{x})^2$
5	5 – 13 = -8	64
9	9 – 13 = -4	16
16	16 – 13 = 3	9
17	17 – 13 = 4	16
18	18 – 13 = 5	25
		$\sum(x-\bar{x})^2 = 130$

(i) Population Variance :

$\sigma^2 = \frac{\sum(x-\mu)^2}{N} = \frac{130}{5} = 26$

(ii) Sample Variance:

$S^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{130}{4} = 32.5$

Example 2: Calculate Population Standard Deviation and Sample Standard Deviation of the given data:

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

Solution:

Here n = 7

Now $\bar{x} = \frac{\sum x}{n} = \frac{6+2+4+9+1+3+5}{7} = \frac{30}{7} = 4.29$

x	$x - \bar{x}$	$(x - \bar{x})^2$
6	6 – 4.29 = 1.71	2.92
2	2 – 4.29 = - 2.29	5.24
4	4 – 4.29 = -0.29	0.08
9	9 – 4.29 = 4.71	22.18
1	1 – 4.29 = -3.29	10.82
3	3 – 4.29 = -1.29	1.66
5	5 – 4.29 = 0.71	0.50
		$\sum(x - \bar{x})^2 = 43.4$

(i) Population Standard Deviation:

$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}} = \sqrt{\frac{43.4}{7}} = \sqrt{6.2} = 2.48$

(ii) Sample Standard Deviation:

$S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{43.4}{6}} = \sqrt{7.23} = 2.68$

Example 3: Find Sample Variance and Sample Standard Deviation of the given data:

7, 10, 9, 14, 11, 8, 5, 12, 8, 3, 13, 16

Example 4: Find Population Variance and Sample Standard Deviation of the following data:

12, 29, 14, 15, 20, 10, 13, 14, 18, 22, 9

11)Frequency Distribution:

Frequency distribution is useful to grouping data. Frequency distribution is a summary of data presented in the form of class intervals and frequencies.

Important parameters for Frequency Distribution:

(i) Determine range of the data

(ii) Determine number of the classes for the frequency distribution

(iii) Determine class width

Important Formula for Frequency Distribution:

(i) Class Midpoint $= \dfrac{Lower\ limit + Upper\ limit}{2}$

(ii) Class Width $= \dfrac{Maximum\ Value - Minimum\ Value}{n}$ **where n = Number of classes**

(iii) Relative Frequency $= \dfrac{Frequency\ of\ the\ class}{Total\ Frequency}$

(iv) Percentage Relative Frequency $= \dfrac{Frequency\ of\ the\ class}{Total\ Frequency} \times 100$ **%**

Example 1:

Construct the frequency distribution of the following data by using class width 2 . Also find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of the frequency distribution:

1	3	2	4	2	3	1	3
7	7	5	3	2	1	7	1

3	4	7	9	10	7	10	9
3	5	6	7	9	4	5	9
8	7	3	4	2	1	3	7
5	9	7	3	4	3	2	1

Solution:

Class	Frequency Distribution	Frequency
1 – 3	⦀⦀, ⦀⦀⦀, ⦀	11
3 – 5	⦀⦀, ⦀⦀⦀, ⦀⦀⦀	15
5 – 7	⦀⦀⦀	5
7 – 9	⦀⦀, ⦀⦀⦀	10
9 – 11	⦀⦀, ⦀	7
		48

(Note : To construct frequency distribution , don't consider 3 in the 1ˢᵗ class , 5 in the 2ⁿᵈ class and so on…..i.e, end point or upper point is not considerable in that class.)

Class	Frequency	Mid point	Relative Frequency	Percentage Relative	Cumulative Frequency

				Frequency	
$1-3$	11	$\dfrac{Lower\ limit + Upper\ limit}{2}$ $= \dfrac{1+3}{2} = 2$	$\dfrac{Frequency\ of\ Class}{Total\ Frequency}$ $= \dfrac{11}{48}$ $= 0.22$	0.22×100 % $= 22\%$	11
$3-5$	15	$\dfrac{3+5}{2} = 4$	$\dfrac{15}{48} = 0.31$	$0.31 \times$ 100% $= 31$	$11+15 = 26$
$5-7$	5	$\dfrac{5+7}{2} = 6$	$\dfrac{5}{48} = 0.10$	$10\ \%$	$11+15+5 = 31$
$7-9$	10	$\dfrac{7+9}{2} = 8$	$\dfrac{10}{48} = 0.20$	20%	$11+15+5+10$ $=41$
$9-11$	7	$\dfrac{9+11}{2} = 10$	$\dfrac{7}{48} = 0.14$	14%	$11+15+5+10+7=48$

Example 2 :

Construct the frequency distribution of the following data by using class width 3 . Also find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of the frequency distribution:

1	9	2	4	2	3	11	3
17	7	5	3	2	0	7	1
3	4	7	9	10	8	10	9
3	5	6	15	9	14	5	0
8	7	3	4	2	1	3	7
5	9	7	3	4	3	2	1
0	11	14	17	13	16	14	12

17	15	16	14	12	13	10	11

Example 3:

(i) Construct a frequency distribution using 5 class interval. Also find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of each class.

(ii) Construct another frequency distribution using 10 class interval and find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of each class.

42	70	64	47	66	69	73	38	48	25
55	85	10	24	45	31	62	47	63	84
16	40	81	15	35	17	40	36	44	17
38	79	35	36	23	64	75	53	31	60
31	38	52	16	81	12	61	43	30	33

Solution:

(i) Using 5 Class interval

Minimum Value = 10 , Maximum Value = 85

Class Width $= \frac{Maximum\ Value - Minimum\ Value}{2} = \frac{85-10}{5} = \frac{75}{5} = 15$

∴ Class Width (Class Length)= 16

Class	Frequency Distribution	Frequency
10 – 26	‖‖, ‖‖	10
26 – 42	‖‖, ‖‖, ‖‖	14
42 – 58	‖‖, ‖‖	10

58 – 74	∭, ‖‖	**10**
74 – 90	∭,		**6**
		50

| Class	Frequency	Mid point	Relative Frequency	Percentage Relative Frequency	Cumulative Frequency
10 – 26	10	$\dfrac{Lower\ limit + Upper\ limit}{2}$ $= \dfrac{10+26}{2}$ $= 18$	$\dfrac{Frequency\ of\ Class}{Total\ Frequency}$ $= \dfrac{10}{50}$ $= 0.2$	0.2×100 % $= 20\%$	10
26 – 42	14	$\dfrac{26+42}{2} = 34$	$\dfrac{14}{50} = 0.28$	$0.31 \times 100\%$ $=28$	10+14 =24
42 – 58	10	$\dfrac{42+58}{2} = 50$	$\dfrac{10}{50} = 0.2$	20 %	10+14+10 =34
58 – 74	10	$\dfrac{58+74}{2} = 66$	$\dfrac{10}{50} = 0.2$	20%	10+14+10+10 =44
74 – 90	6	$\dfrac{74+90}{2} = 82$	$\dfrac{6}{50} = 0.12$	12%	10+14+10+10+6 =50

(ii) Use 10 class interval

Minimum Value = 10 , Maximum Value = 85

Class Width $= \dfrac{Maximum\ Value - Minimum\ Value}{2} = \dfrac{85-10}{10} = \dfrac{75}{10} = 7.5$

∴ Class Width (Class Length)= 8

Class	Frequency Distribution	Frequency
10 – 18	‖‖, ‖	7
18 – 26	‖‖	3
26 – 34	‖‖	5
34 – 42	‖‖, ‖‖	9
42 – 50	‖‖, ‖	7
50 – 58	‖‖	3
58 – 66	‖‖, ‖	6
66 – 74	‖‖	4
74 – 82	‖‖	4
82 – 90	‖	2
		50

Example 4:

Construct a frequency distribution using 4 class interval. Also find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of each class.

0	2	4	6	8	10	12	14	1
3	5	7	9	11	13	15	2	4
6	8	10	12	14	15	13	11	9
7	5	3	1	0	4	8	12	3
6	9	12	1	7	9	11	2	8

Example 1:

Construct the frequency distribution of the following data by using class width 3 . Also find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of the frequency distribution:

1	3	2	4	2	3	1	3
7	7	5	3	2	1	7	1
3	4	7	9	10	7	10	9
3	5	6	7	9	4	5	9
8	7	3	4	2	1	3	7
5	9	7	3	4	3	2	1

Example 2 :

Construct the frequency distribution of the following data by using 4 class . Also find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of the frequency distribution:

1	9	2	4	2	3	11	3
17	7	5	3	2	0	7	1

3	4	7	9	10	8	10	9
3	5	6	15	9	14	5	0
8	7	3	4	2	1	3	7
5	9	7	3	4	3	2	1
0	11	14	17	13	16	14	12
17	15	16	14	12	13	10	11

Example 3:

Construct a frequency distribution using 4 class interval. Also find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of each class.

0	2	4	6	8	10	12	14	1
3	5	7	9	11	13	15	2	4
6	8	10	12	14	15	13	11	9
7	5	3	1	0	4	8	12	3
6	9	12	1	7	9	11	2	8

Example 4:

Construct a frequency distribution using class width 5. Also find mid point of each class, relative frequency, percentage relative frequency and cumulative frequency of each class.

0	24	2	16	18	20	22	24	3
1	23	4	6	9	12	15	18	21
2	22	6	4	8	12	16	20	24
3	21	8	5	10	15	20	6	12
4	20	10	18	7	14	21	8	16

5	19	12	11	13	15	17	19	5

12) Data Graphs:

There are two types of Data Graphs:

(i) Quantitative Data Graphs:

(a) Histogram (b) Frequency Polygon (c) Ogive
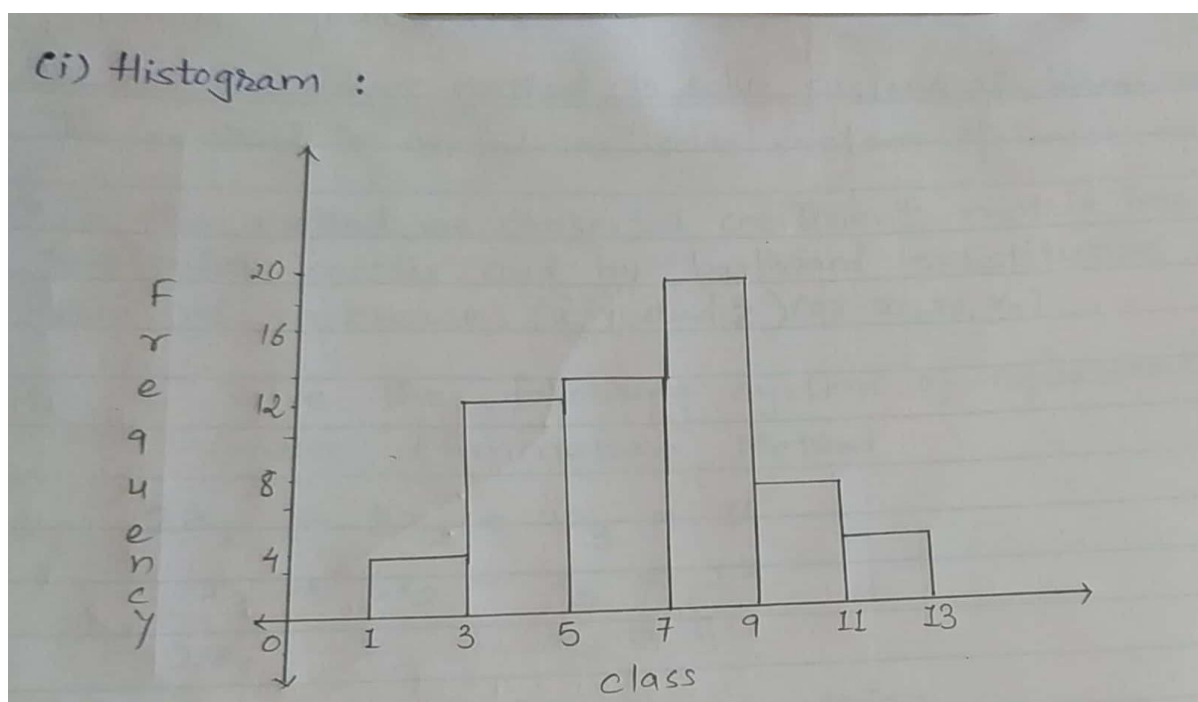
(ii) Qualitative Data Graphs :

(a) Bar Graph (b) Pie Chart

(i) Quantitative Data Graphs:

Example 1: Construct Histogram , Frequency Polygon and Ogive of the following data:

Class	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13
Frequency	4	12	13	19	7	5

(a) Histogram (X – axis consider class point , Y – axis consider Frequency)

(b) Frequency Polygon: (X – axis consider mid point of each class , Y – axis consider Frequency)

Class	Frequency	Mid Point
1 – 3	4	$\frac{1+3}{2} = 2$
3 – 5	12	4
5 – 7	13	6
7 – 9	19	8
9 – 11	7	10
11 – 13	5	12

(c) Ogive : (X – axis consider end point of each class , Y – axis consider Cumulative Frequency)

Class	Frequency	Cumulative Frequency
1 – 3	4	4
3 – 5	12	4+12 = 16
5 – 7	13	4+12+13 = 29
7 – 9	19	4+12+13+19 = 48
9 – 11	7	4+12+13+19+7 =55
11 – 13	5	4+12+13+19+7+5 = 60

(iii) Ogive :-
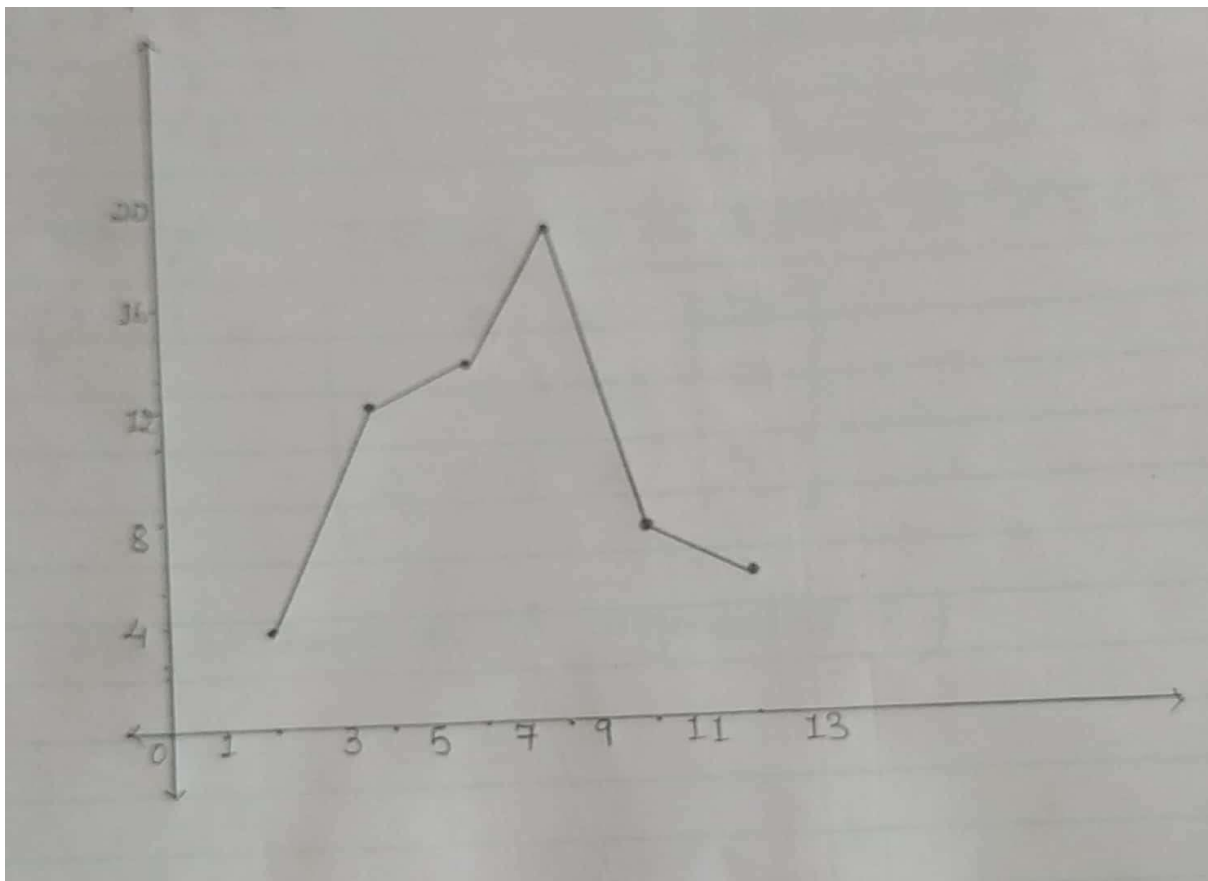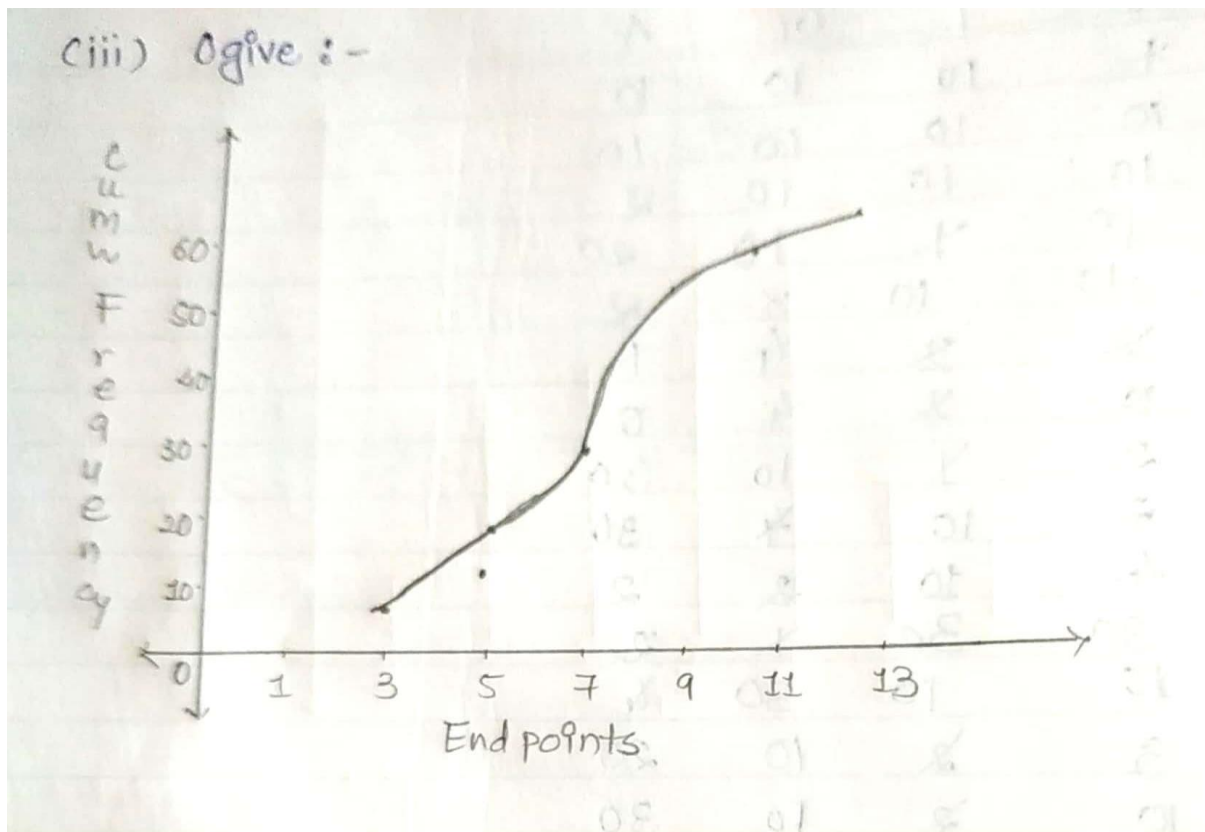
Scanned with CamScanner

Example 2 : Construct Histogram , Frequency Polygon and Ogive of the following data:

Class	1 – 4	4 – 7	7 – 10	10 – 13	13 – 16
Frequency	5	11	9	15	20

Example 3: Construct Frequency Distribution of the given data using class width 3 and construct histogram and ogive for that

0	2	4	6	8	10	12	14
1	3	5	7	9	11	13	3
6	9	12	13	4	8	12	6
12	7	14	8	16	10	11	12
0	1	2	3	4	5	6	7

Example 4 : Construct Frequency distribution of the given data using 4 class and construct frequency polygon and ogive

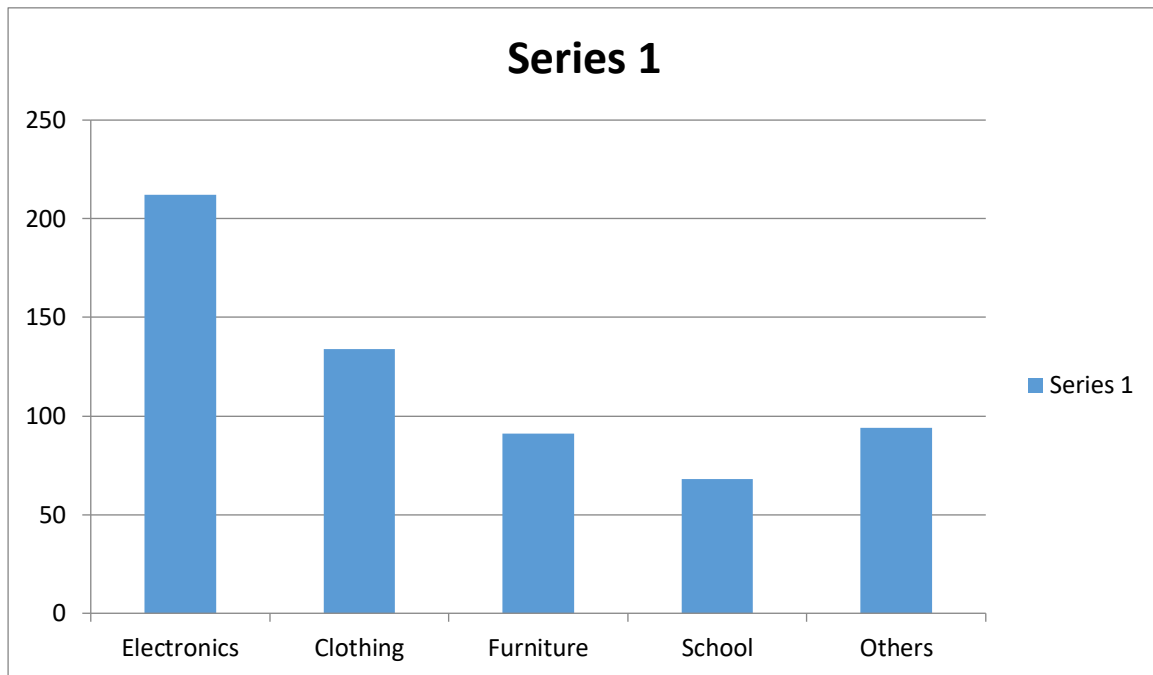| 1	16	2	15	3	14	4	13	5	12
6	11	7	10	8	9	2	4	6	8
10	12	14	16	1	3	5	7	9	11
13	15	2	10	12	1	11	13	14	15
9	6	8	7	4	5	6	3	2	12

BAR GRAPH

Example 1: Construct the Bar Plot (Bar Graph) of the following data:

Electronics	Clothing	Furniture	School	Others
211.89	134.40	90.90	68.47	93.72

Example 2: Construct the bar graph of the following data

A	B	C	D	E	F	G	H
10.5	29.8	13.4	15.8	35.7	44.8	57	63
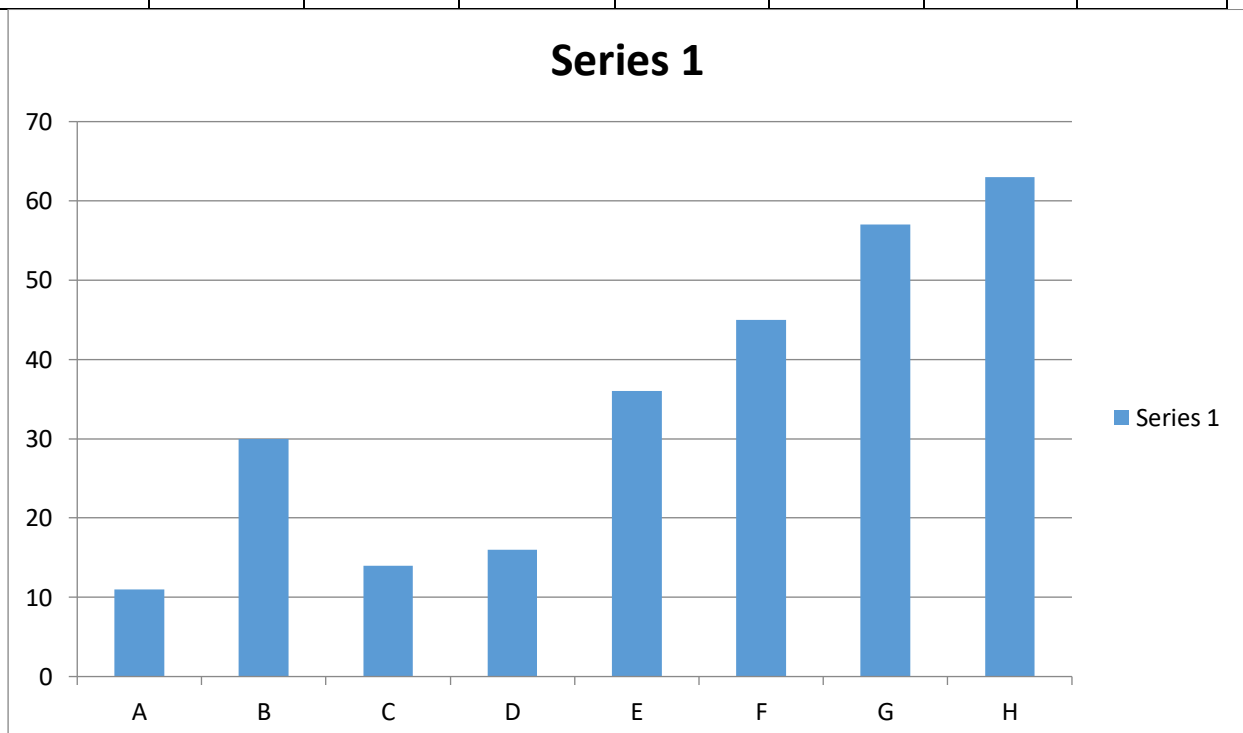
13) Stem and Leaf Display:

By default Stem contains 2 digits and leaf contains 1 digit (when data contains 3 digits)

In stem repetition is not allowed, in leaf repetition is allowed.

Stem and Leaf display :

①

→ By default stem

* Steam and Leaf display

Stem 2 ↙ 3 ↘ Leaf
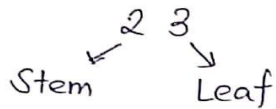
Example 1 : Construct the stem and leaf display of the following data :

23	55	87	86	78	23	23	23	27
81	10	73	79	81	69	68	73	83
41	33	80						

Solⁿ Step 1 :

                    Leaf

Stem
```
2 | 3  3 3 3  7
5 | 5
8 | 7  6  1  1   3  0
7 | 8  3  9  3
1 | 0
6 | 9  8
4 | 1
3 | 3
```

Step 2: Ascending Order in Stem and leaf   ②

	Leaf
S 1	0
t 2	3 3 3 7
e 3	3
m 4	1
5	5
6	8 9
7	3 3 8 9
8	0 1 1 3 6 7

Example : 2 Construct the stem and Leaf display
of the given data :

212	234	219	210	700	259	257	249
220	249	243	212	275	243	218	243
219	217	273	247	211	214		

Step 1

Stem	Leaf
21	2 9 0 2 8 9 7 1 4
23	4
70	0
25	9 7
24	9 9 3 3 3 7
22	0
27	5 3

S t e m

Step 2   Ascending Order

Stem	Leaf
21	0 1 2 2 4 7 8 9 9
22	0
23	4
24	3 3 3 7 9 9
25	7 9
27	3 5
70	0

Example : Obtain Stem and Leaf Display of the given data:

(1)

23	24	46	78	96	55	45	32
31	65	98	74	12	13	22	23
33	36	44	45	46	47	48	49
55	57	66	69	77	78	80	81
91	92	32	35	36	39	74	78
64	65	20	21	23	66	31	10

(2)

112	113	118	123	124	126
128	129	131	132	144	148
201	204	253	258	256	301
303	312	314	358	359	412
414	423	456	501	505	458
630	635	141	123	201	206
231	235	309	789	741	203
697	214	259	364	201	104

Multiple Choice Questions:

(1) What is the length of the class 1 – 4?

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

(2) What is the value of relative frequency if total frequency is 50 and frequency of the class is 17?

(a) 17 (b) 0.17 (c) 0.34 (d) 50

(3) What is the mean of the data 0, 8, 9, 7, 4, 3?

(a) 3.789 **(b) 5. 1666** (c) 0 (d) 5. 1004

(4) What is the mean of first ten natural numbers?

(a) 5 (b) 55 **(c) 5.5** (d) 0

(5) What is the value of P_{30} for the data 2, 7, 9, 11, 14, 15, 19?

(a) 19 (b) 15 (c) 11 **(d) 9**

N= 7

(6) Which of the following is not a parameter of five number summary?

(a) Median (b) First Quartile **(c) Mean** (d) Third Quartile

(7) Q_3 = _____

(a) P_{25} (b) P_{50} **(c) P_{75}** (d) P_{100}

(8) What is Q_4 of the data 90, 89, 101, 67, 89, 25, 87, 94, 98, 88, 81, 57, 77?

(a) 89 (b) 98 **(c) 101** (d) 57

(9) What is mode of the data 0, 9, 8, 7, 5, 6, 4, 3, 2, 1, 9?

(a) 5. 89 (b) 0 **(c) 9** (d) 8

(10) What is Range of 5, 9, 2, 7, 11, 12, 14, 7, 15, 13, 10?

a) 5 b) 10 c) 7 **d) 13**

(11) What is Mean of the data 1, 7, 3, 0, 4, 2, 9, 5 ?

a) 3.875 b) 0 c) 3 d) 2.915

(12) What is First Quartile of the data 11, 15, 17, 19, 24, 25, 32 ?

a) 17 b) 11 **c) 15** d) 19

(13) If observations are not repeated then Z = ____

a) $2M - 3\overline{x}$ b) $2M + \overline{x}$ c) $3M + 2\overline{x}$ **d) $3M - 2\overline{x}$**

(14) What is Median of 1, 2, 4, 6, 7, 8, 10, 11, 14, 15?

a) 7.5 b) 9 c) 7 d) 8.5

(15) What is P_{30} of 10, 19, 23, 24, 37, 56, 69, 78, 100?

a) 23 b) 10 c) 19 d) 24

(16) Which of the following is the correct formula of MAD?

a) $\frac{\sum(x-\bar{x})}{n}$ b) $\frac{\sum(x-\bar{x})}{n-1}$ c) $\frac{\sum|x-\bar{x}|}{n}$ d) $\frac{\sum|x-\bar{x}|}{n-1}$

(17) Which of the following is not a quantitative graph?

a) Histogram b) Frequency polygon c) Ogive (d) Bar Graph

(18) What is Q_3 of 2, 4, 6, 8, 13, 15, 19, 21?

a) 17 b) 12 c) 15 d) 21

(19) What is the value of sample standard deviation ,if sample variance has value 49?

a) 48 b) 7 c) 6 d) 2401

(20) What is the value of population standard deviation, if population variance has value 225?

a) 15 b) 224 c) 625 d) 50625

(21) Consider the following frequency distribution:

Class	Frequency
1 – 4	3
4 – 7	9
7 – 10	4
10 – 13	5
13 – 16	8

What is cumulative frequency of class 7 – 10 ?

a) 12 b) 4 c) 17 d) 16

(22) Consider the following frequency distribution:

Class	Frequency
0 – 4	4

4 – 8	10
8 – 12	14
12 – 16	8
16 – 20	14

What is percentage frequency of class 4 – 8?

a) 40% b) 20 % c) 10 % d) 50 %

(23) In ogive Y –axis represents _____.

a) Cumulative Frequency b) Frequency c) Class d) Percentage Frequency

(24) What is Q_2 of the data 55, 100, 125, 140 and 60?

a) 140 b) 154 c) 96 d) 100

(25) Which of the following is correct formula of IQR?

a) $Q_1 – Q_3$ b) $Q_2 – Q_3$ c) $Q_3 – Q_1$ d) $Q_4 – Q_3$