UNIT 11 Information Handling

Students' Learning Outcomes

After completing this unit, the students will be able to:

- Construct cumulative frequency table and cumulative frequency polygon or Ogive.
- Interpret the median, quartiles, deciles, percentiles and inter quartile range from cumulative frequency curve.
- Interpret and analyze box and whisker plots.
- Construct and interpret data from scatter diagrams and also draw lines of best fit.
- Measure correlation using scatter diagram.
- Calculate the range, standard deviation and variance for grouped data.
- Use the mean and standard deviation to compare two sets of data.
- Solve real life situations involving variance and standard deviation for grouped data.
- Apply concepts from measures of dispersion to solve real life situations (determining the consistency of data, checking variability in forecasting, manufacturing, finance and economics).

11.1 Construction of Cumulative Frequencies

Cumulative frequency is the running total of the frequencies, where each frequency is added to the sum of all frequencies before it.

A table of cumulative frequencies, also known as a cumulative frequencies distribution, presents the running total of frequencies for each data point or class interval showing how many observations fall at or below a specific value.

Example 1: The marks of 50 college students out of 100 are given below. Construct cumulative frequency table.

85	66	76	45	66	91	77	64	71	74
47	78	76	42	70	58	71	67	80	78
73	48	68	87	81	72	65	69	78	84
75	53	58	87	56	72	62	92	73	83
97	81	51	61	58	72	62	79	88	74

Solution : Highest marks = 97, Lowest marks = 42

The same 50 scores grouped into a frequency distribution.

Table 11.1 Cumulative Frequency Table

Class Intervals	Tally Marks	Frequency	Cumulative Frequencies
40 - 44	|	1	1
45 - 49	|||	3	1 + 3 = 4
50 - 54	||	2	4 + 2 = 6
55 - 59	||||	4	6 + 4 = 10

60 - 64	𝍷𝍷𝍷𝍷	4	10 + 4 = 14
65 - 69	𝍸𝍷	6	14 + 6 = 20
70 - 74	𝍸𝍸	10	20 + 10 = 30
75 - 79	𝍸𝍷𝍷𝍷	8	30 + 8 = 38
80 - 84	𝍸	5	38 + 5 = 43
85 - 89	𝍷𝍷𝍷𝍷	4	43 + 4 = 47
90 - 94	𝍷𝍷	2	47 + 2 = 49
95 - 99	𝍷	1	49 + 1 = 50
Total		$n = \sum f = 50$	

Cumulative Frequency Polygon or Ogive

A cumulative frequency polygon, also known as an ogive, is a graphical representation that shows the cumulative total of frequencies up to each class boundary in a frequency distribution.

There are two methods for construction of a cumulative frequency polygon or ogive:

(a) Less than method **(b)** More than method

Steps for Construction of Frequency Polygon or Ogive

Step I: Prepare a cumulative frequency table from the given data.

Step II: Mark the upper class boundaries on the x-axis for less than method.
Mark the lower class boundaries on the x-axis for more than method.

Step III: Mark the corresponding cumulative frequencies on the y-axis.

Step IV: Plot each point using the upper class boundary or lower class boundary on the x-axis and its cumulative frequency on the y-axis.

Step V: Connect the plotted points using a smooth freehand curve or use straight line segments to form the cumulative frequency polygon or ogive.

Important Definitions

Median:

The median is the middle value in a set of data when the values are arranged in ascending or descending order. It divides the data into two equal parts.

Quartiles:

Quartiles divide an arranged data set into four equal parts, each containing 25% of the data.

Q_1 (First Quartile)

25% of the data falls below this value.

Q_2 (Second Quartile)

50% of the data falls below this value (also called the median).

Q_3 (Third Quartile)

75% of the data falls below this value.

> **Remember:**
> Q_1 is the lower quartile, Q_2 is the median and Q_3 is the upper quartile.

Deciles:

Deciles divide an arranged data set into ten equal parts, each representing 10% of the data. In other words, there are nine decile values from D_1 to D_9.

> **Remember!**
> P_{25} is the same as Q_1, P_{50} is the same as the median (Q_2), and P_{75} is the same as Q_3.

For example, D_3 is the value below which 30% of the data lies and D_7 is the value below which 70% of the data falls.

Percentiles:

Percentiles divide an arranged data set into 100 equal parts, each containing 1% of the data. There are ninety nine percentile values from P_1 to P_{99}.

For example, P_{90} means 90% of the data lies below that value.

Interquartile Range (*IQR*)

The interquartile range is a measure of spread that represents the distance between the first quartile Q_1 and the third quartile Q_3. It is calculated as $Q_3 - Q_1$

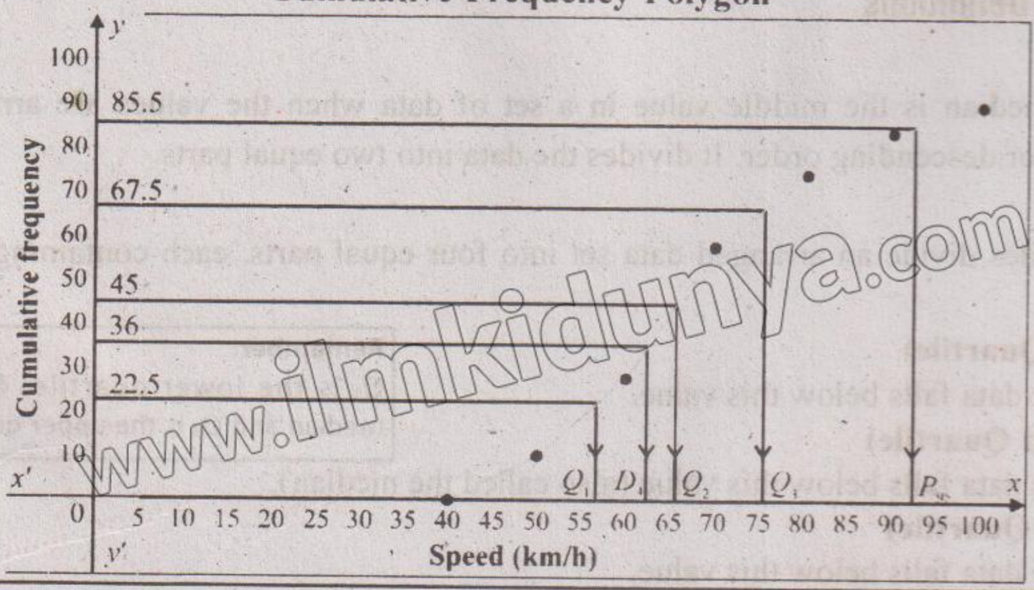
Example 2: The table provides the information about the speeds, in km/h of 90 vehicles passing a speed checkpoint. Construct the cumulative frequency polygon by less than method and interpret median, Q_1, Q_3, D_4, P_{95} and IQR.

Speed (km/h)	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
f	10	18	30	17	9	6

Solution:

Speed (km/h)	f	cf
40 – 50	10	10
50 – 60	18	10 + 18 = 28
60 – 70	30	28 + 30 = 58
70 – 80	17	58 + 17 = 75
80 – 90	9	75 + 9 = 84
90 – 100	6	84 + 6 = 90
Total	$n = \sum f = 90$	

Cumulative Frequency Polygon



The position of median is $\frac{n}{2} = \frac{90}{2} = 45$. Draw a horizontal line from 45 on the y-axis until it meets the curve and then draw vertical line from this point to x-axis. Now, take the reading on the x-axis where the vertical line hits the x-axis. So,

$$\text{Median} = 65.7$$

By the same way, we can find Q_1, Q_3, D_4, P_{95} and semi interquartile range.

The position of Q_1 is, $\frac{n}{4} = \frac{90}{4} = 22.5$. So, from the curve $Q_1 = 56.9$.

The position of Q_3 is $\frac{3n}{4} = \frac{3(90)}{4} = 67.5$. So, from the curve Q_3, $= 75.6$.

The position of D_4 is $\frac{4n}{10} = 36$. So, from the curve $Q_4 = 62.7$.

The position of P_{95} is $\frac{95n}{100} = \frac{95(90)}{100} = 85.5$. So, from the curve $= 92.5$.

$$\text{IQR} = Q_3 - Q_1 = 75.6 - 56.9 = 18.7$$

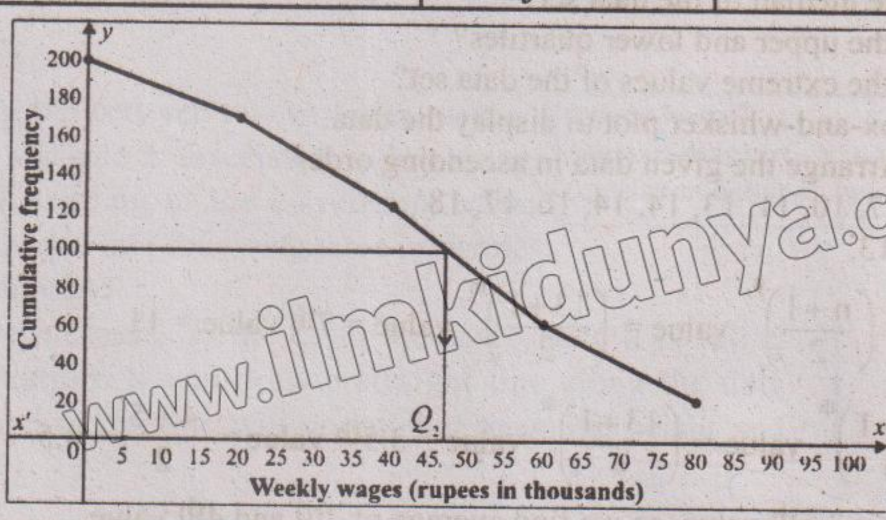
Example 3: Draw the cumulative frequency polygon for the following frequency distribution of the weekly wages (rupees in thousands) of number of workers by more than method.

Weekly wages (rupees in thousands)	0–20	20–40	40–60	60–80	80–100
No. of workers	30	48	61	40	21

Also interpret Q_2.

Solution:

Weekly wages (rupees in thousands)	No. of workers (f)	cf
0–20	30	200
20–40	48	200–30= 170
40–60	61	170–48= 122
60–80	40	122–61 =61
80–100	21	61–40 = 21
Total	$\sum f = 200$	



The position of Q_2 is $\frac{n}{2} = \frac{200}{2} = 100$. So, from the curve $Q_2 = 47$

11.2 Box-and-Whisker Plot

A box-and-whisker plot or box plot is a graphical representation of a data set five-number summary (minimum, first quartile, median, third quartile and maximum). It is useful for visualizing data distribution and comparing data sets.

There are following key components of box plot:

The Box: It represents the interquartile range, which is middle 50% of the data with the box edges marking the first Q_1 and third Q_3 quartiles.

The median: A line within the box representing the second quartile (Q_2) or the middle of the data set.

Whiskers: Lines extending from the box to the minimum and maximum values, showing the overall spread of the data.

We can show this information on a box-and-whisker plot. To begin with, we draw a horizontal number line using a suitable scale. On the number line, the minimum score, the maximum score and the quartiles are indicated as shown in the figure given below:

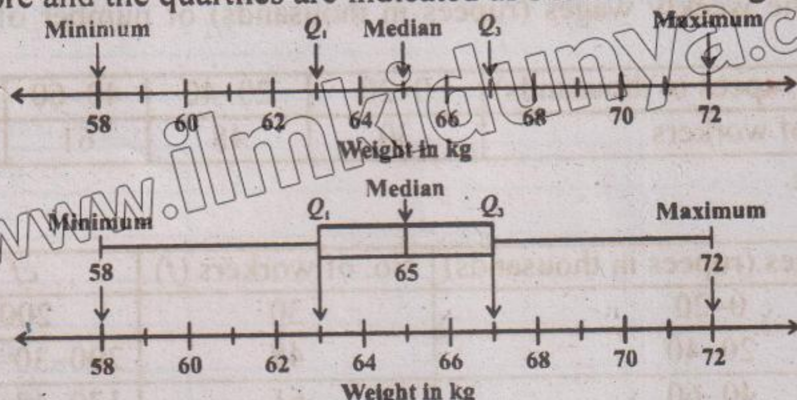


Figure 11.3

Example 4: For the given set of data,

2, 10, 3, 16, 14, 13, 4, 7, 11, 5, 17, 14, 18

(i) What is the median of the data set?

(ii) What are the upper and lower quartiles?

(iii) What are the extreme values of the data set?

(iv) Draw a box-and-whisker plot to display the data.

Solution: (i) Arrange the given data in ascending order.

2, 3, 4, 5, 7, 10, 11, 13, 14, 14, 16, 17, 18

Here $n = 13$,

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th} \text{value} = \left(\frac{13+1}{2}\right)^{th} \text{value} = 7^{th} \text{ value} = 11$$

(ii) $$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{value} = \left(\frac{13+1}{4}\right)^{th} \text{value} = 3.5^{th} \text{ value} = \frac{4+5}{2} = 4.5$$

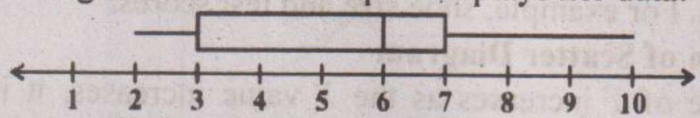
As the Q_1 is 3.5^{th} value, so we find average of 3^{rd} and 4^{th} value.

(iii) Minimum value = 2, Maximum value = 18

(iv) The box-and whisker plot is drawn below:

Fig. 11.4

Example 5: Kiran recorded the number of days it rained or snowed each month during the last year. The following box-and-whisker plot displays her data:



(i) What do the extremes tell about the number of days it rained or snowed?

(ii) What is the median number of days it rained or snowed?

(iii) What are the upper and lower quartiles for the data?

(iv) Where the most data is clustered?

Solution

(i) The least number of days it rained or snowed in a month was 2 and the greatest number of days it rained or snowed in a month was 10.

(ii) Median = 6 days

(iii) Lower quartile = 3 days, Upper quartile = 7 days

(iv) Most of the data is clustered in the box.

11.3 Scatter Diagram

A scatter diagram (or scatter plot) is a graphical tool used to display the relationship between two numerical variables. Each point on the graph represents a pair of values. One variable is plotted on the horizontal axis and the other is plotted on the vertical axis.

Line of Best Fit

A line of best fit (also called a trend line) is a straight line drawn through a scatter diagram that best represents the overall direction of the data.

It is called an estimated line of best fit because it is drawn manually, based on general guidelines to best represent the overall trend of the data. It should go through the middle of the points. Ideally, there should be about the same number of points above and below the line.

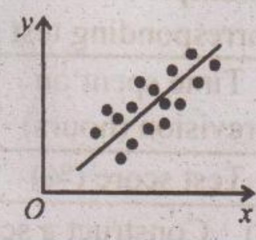
Correlation:

It is the relationship between two variables, where a change in one variable is associated with a change in the other. According to the correlation, scatter diagrams are divided into following three categories:

> **Do you know?**
> Scatter diagram visually shows the correlation between two variables.

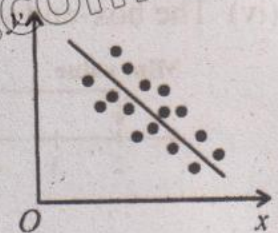
Positive Correlation:

If one variable increases as the other increases, then it is called positive correlation. If you draw a straight line along the data points, the slope of the line will go up. For example, height and weight.



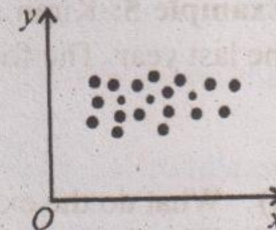
Negative Correlation

If one variable increases as the other decreases, then it is called negative correlation. If you draw a straight line along the data points, the slope of the line will go down. For example, speed and travel time.



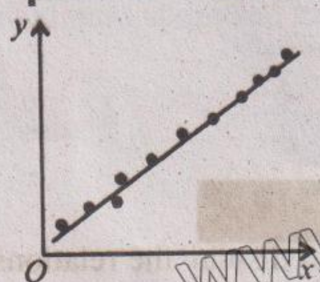
Zero Correlation

If there is not consistent pattern between the two variables, then it is called no correlation. For example, shoe size and test scores.



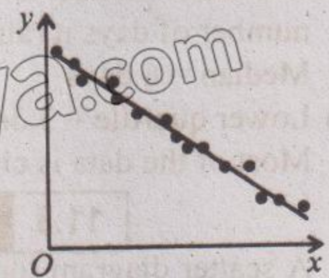
Correlation Pattern of Scatter Diagram

- When the value of Y increases as the X value increases, it is called **strong positive correlation pattern.**
- When the value of Y decreases as the X value increases, it is called strong **negative correlation pattern.**
- When the value of Y increases slightly as the X value increases, it is called **weak positive correlation pattern.**



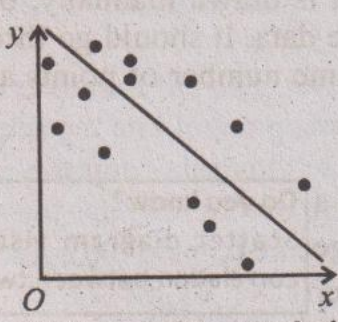
Strong positive correlation

Weak positive correlation

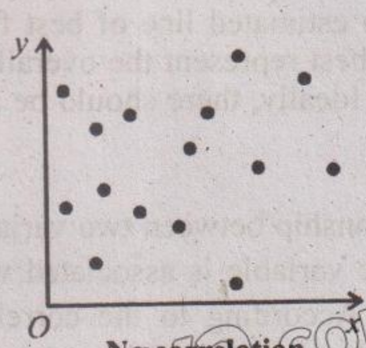


Strong negative correlation

- When the value of Y decreases slightly as the X value increases, it is called **weak negative correlation pattern.**
- When there is no connection between the two variables, it is called **no-correlation pattern.**



Weak negative correlation



No correlation

Example 6: The table below displays the amount of time spent on revision and the corresponding test scores for ten students.

Time spent on revision (hours)	0.5	1	4	6	2	3	7	5	8	9
Test score (%)	22	38	62	68	30	46	70	60	86	90

(i) Construct a scatter diagram and draw a line of best fit on it.

(ii) Describe the correlation shown in the scatter diagram.
(iii) Saleem spent 5.5 hours on revision. Use line of best fit to estimate his test score.

Solution:

(i)

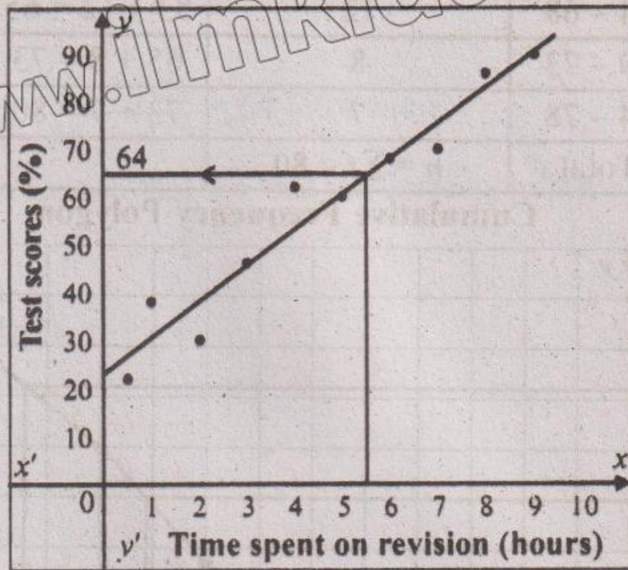


Fig. 11.5

Steps:

- Draw x-axis for "Time spent on revision" and y-axis for "Test scores".
- For each pair from the table, mark a point on the graph.
- Observe the pattern of points and draw a line that should be close as possible to all points.

(ii) When the time spent on revision increases, the test score increases. Therefore, strong positive correlation is between given two variables.
(iii) At 5.5 hours, test score of Saleem will be 64%.

Solved Exercise 11.1

1. The following distribution represents the achievement of a student's group upon a memory test:

(i) Plot cumulative frequency graph of group scores.

(ii) Determine median, lower and upper quartiles, D_3, P_{90} and interquartile range graphically.

Scores	44–48	49–53	54–58	59–63	64–68	69–73	74–78
No. of Students	3	12	15	23	12	8	7

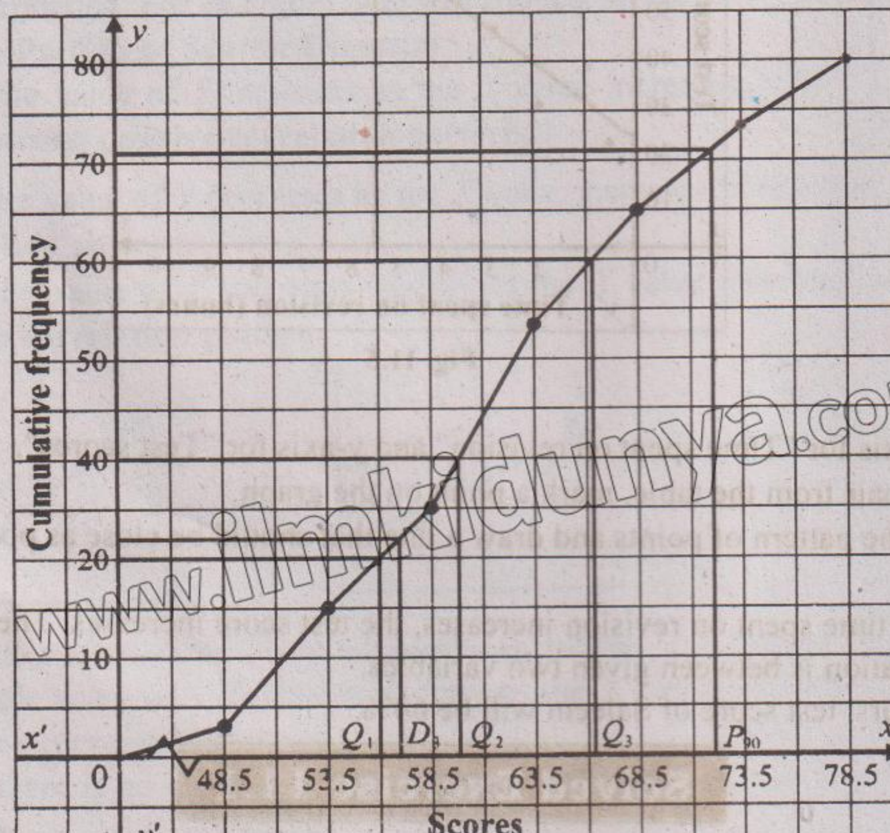
(i) Plot cumulative frequency graph of group scores.

Sol. **Cumulative Frequency Polygon**

Scores	No. of Students f	c.f.
44 – 48	3	3
49 – 53	12	3 + 12 = 15

54 – 58	15	15 + 15 = 30
59 – 63	23	30 + 23 = 53
64 – 68	12	53 + 12 = 65
69 – 73	8	65 + 8 = 73
74 – 78	7	73 + 7 = 80
Total	$n = \sum f = 80$	

Cumulative Frequency Polygon



(ii) Determine median, lower and upper quartiles, D_3, P_{90} and interquartile range graphically.

Scores	44–48	49–53	54–58	59–63	64–68	69–73	74–78
No. of Students	3	12	15	23	12	8	7

Sol. Median The position of median is $\frac{n}{2} = \frac{80}{2} = 40$

Draw a horizontal line from 40 on the y-axis until it meets the curve and then draw vertical line from this point to x-axis.

Now, take the reading on the x-axis, where the vertical line hits the x-axis. So

$\Rightarrow$ Median = 61

Lower Quartile (Q_1)

The position of Q_1 is $\frac{n}{4} = \frac{80}{4} = 20$ (20th value)

Now from the graph we see that the corresponding value of 20 on Y-axis is 55 on X-axis

$\Rightarrow Q_1 = 55$

Upper Quartile (Q_3)

The position of Q_3 is $\frac{3n}{4} = \frac{3(80)}{4} = 60$ (60th value)

Now from the graph we see that see that the corresponding value of 60 on Y-axis is 66 on X-axis.

$\Rightarrow Q_3 = 66$

Interquartile Range (IQR)

$\therefore \text{IQR} = Q_3 - Q_1$

$= 66 - 55$

$= \text{IQR} = 11$

3rd Decile (D_3)

The position of D_3 is $\frac{3n}{10} = \frac{3(80)}{10} = 24$ (24th value)

Now fromt he graph we see that the corresponding value of 24 on Y-axis is 57 on X-axis.

$\Rightarrow Q_3 = 57$

90th Percentile (P_{90})

The position of P_{90} is $\frac{90n}{100} = \frac{90(80)}{100} = 72$ (72th value)

Now from the graph we see that the corresponding value of 72 on Y-axis is 73 on X-axis. $\Rightarrow P_{90} = 73$

2. Construct an ogive for the following distribution of scores:

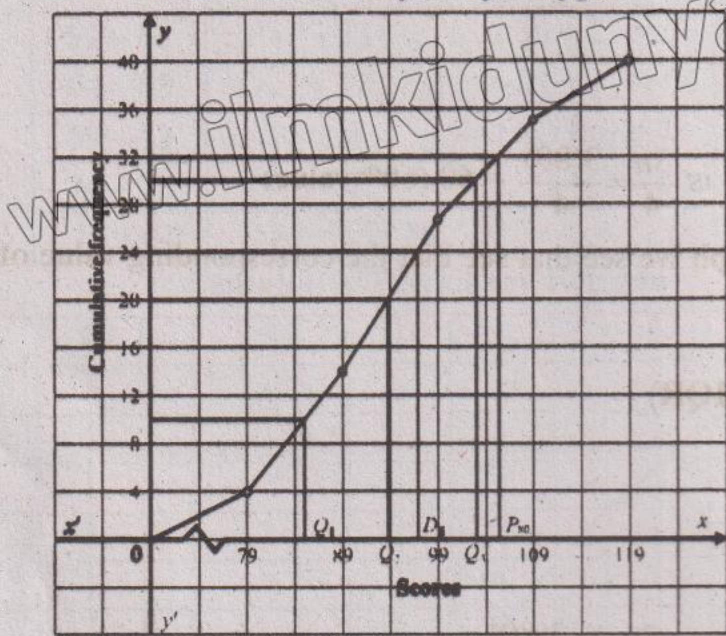
Scores	69–79	79–89	89–99	99–109	109–119
f	4	10	13	8	5

Determine median, lower and upper quartiles, D_6, P_{80} and interquartile range graphically.

Sol.

Scores	*f*	*cf*
69 – 79	4	4
79 – 89	10	4 + 10 = 14
89 – 99	13	14 + 13 = 27
99 – 109	8	27 + 8 = 35
109 – 119	5	35 + 5 = 40
Total	$n-\sum f = 40$	

Cumulative Frequency Polygon



Median: The position of median is $\frac{n}{2} = \frac{40}{2} = 20$ (20th value)

Draw a horizontal line from 20 on Y-axis until it meets the curve and then draw vertical line from this point to X-axis. So,

Median = 94

Lower Quartile (Q_1) : The position of Q_1 is $\frac{n}{4} = \frac{40}{4} = 10$ (10th Value)

Now from the graph we see that the corresponding value of 10 on Y-axis is 85 on X-axis.

$\Rightarrow Q_1 = 85$

Upper Quartile (Q_3) The position of Q_3 is $\frac{3n}{4} = \frac{3(40)}{4}$

$= 30$ (30th value)

Now from the graph we see that the corresponding value of 30 on Y-axis is 103 on X-axis.

$\Rightarrow Q_3 = 103$

6th Decile (D_6) The position of D_6 is $\frac{6n}{10} = \frac{6(40)}{10} = 24$ (24th value)

Now from the graph we see that the corresponding value of 24 on Y-axis is 97 on X-axis.

$\Rightarrow D_6 = 97$

80th Percentile (P_{80}) The position of P_{80} is $\frac{80n}{100} = \frac{80(40)}{100} = 32$ (32th value)

Now from the graph we see that the corresponding value of 32 on Y-axis is 105 on

X-axis.

$\Rightarrow P_{80} = 105$

Interquartile Range (IQR)

$$\therefore IQR = Q_3 - Q_1$$
$$= 103 - 85$$
$$IQR = 18$$

3. Find Q_1, Q_3 median, range and *IQR* for the dataset given below:
3, 6, 8, 4, 7, 5, 10, 11, 13, 9, 14, 12, 15, 16, 17
Draw a box-and-whisker plot.

Sol. Given data in ascending order:
3,4,5,6,7,8,9,10,11,12,13,14,15,16,17
Here
Total No, of values = n = 15

Median:

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th} \text{Value}$$
$$= \left(\frac{15+1}{2}\right)^{th} \text{Value}$$
$$= \left(\frac{16}{2}\right)^{th} \text{Value}$$
$$= 8\text{th Value} \Rightarrow \text{Median} = 10$$

Lower Quartile (Q_1)

$$Q_1 = \left(\frac{n+1}{4}\right)^{th}$$
$$= \left(\frac{15+1}{4}\right)^{th} \text{Value}$$
$$= \left(\frac{16}{4}\right)^{th} \text{Value}$$
$$= 4\text{th Value} \Rightarrow Q_1 = 6$$

Upper Quartile (Q_3)

$$Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{Value}$$
$$= 3\left(\frac{15+1}{4}\right)^{th} \text{Value}$$
$$= 3\left(\frac{16}{4}\right)^{th} \text{Value}$$
$$= [3(4)]_{th} \text{ Value}$$
$$= 12_{th} \text{ Value} \Rightarrow Q_3 = 14$$

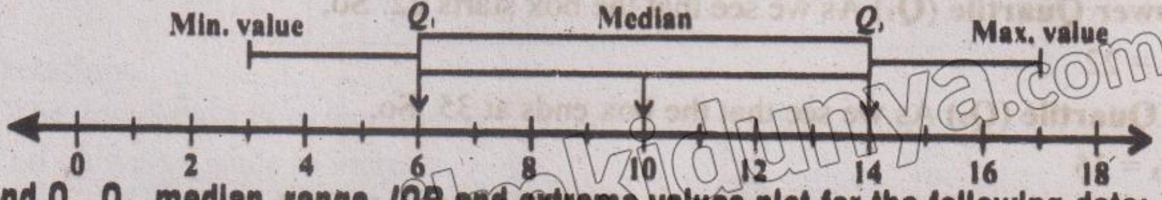
Range

Range = highest value – lowest value
Range = 17 – 3 = 14

Interquartile Range (IQR)

$$IQR = Q_3 - Q_1$$
$$= 14 - 6$$
$$IQR = 8$$

Box – and – Whisker Plot



4. Find Q_1, Q_3, median, range, *IQR* and extreme values plot for the following data:
102, 98, 95, 100, 93, 110, 108, 104, 97, 96, 92, 101, 99, 105, 107
Draw a box-and-whisker.

Sol. Given data set in ascending order
92, 93, 95, 96, 97, 98, 99, 100, 101, 102, 104, 105, 107, 108, 110

Here

Total no. of values = n = 15

Median

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

$$= \left(\frac{15+1}{2}\right)^{th} \text{Value}$$

$$= \left(\frac{16}{2}\right)^{th} \text{Value}$$

$= 8\text{th Value} \Rightarrow \text{Median} = 100$

Lower Quartile (Q_1)

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{Value}$$

$$= \left(\frac{15+1}{4}\right)^{th} \text{Value}$$

$$= \left(\frac{16}{4}\right)^{th} \text{Value}$$

$= 4\text{th Value} \Rightarrow Q_1 = 96$

Upper Quartile (Q_3)

$$Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{Value}$$

$$= 3\left(\frac{15+1}{4}\right)^{th} \text{Value}$$

$$= 3\left(\frac{16}{4}\right)^{th} \text{Value}$$

$= [3(4)]^{th}$ Value

$= 12^{th} \text{ Value} \Rightarrow Q_3 = 105$

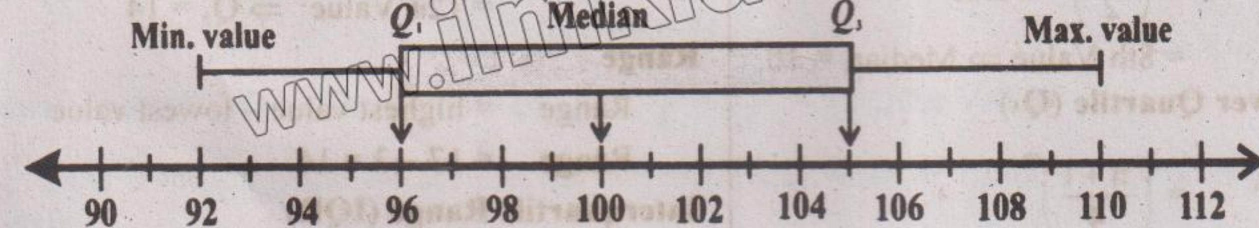
Range

Range = Highest value – lowest value

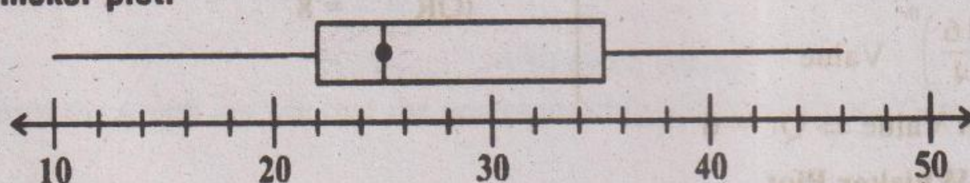
= 110 – 92

$\Rightarrow$ Range = 18

Box-and-whisker plot



5. Find Q_1, Q_3, median, minimum and maximum values for the following box-and-whisker plot:



Sol. Lower Quartile (Q_1) As we see that the box starts 22. So,

$Q_1 = 22$

Upper Quartile (Q_3) As we see that the box ends at 35. So,

$Q_3 = 35$

Median The vertical line indicated by a dot inside the box is at 25. So,

Median = 25

Minimum Value The left Whisker ends exactly at 10. So,

Minimum Value = 10

Maximum Value The right whisker ends exactly at 46. So,

Maximum Value = 46

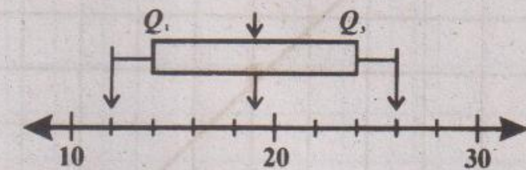
6. Plot the box-and-whisker plot, if min. value = 12, max. value = 26, median = 19, $Q_1 = 14$ and $Q_3 = 24$.

Sol. Given that

min. Value = 12, max. value = 26

median = 19, $Q_1 = 14$ and $Q_3 = 24$

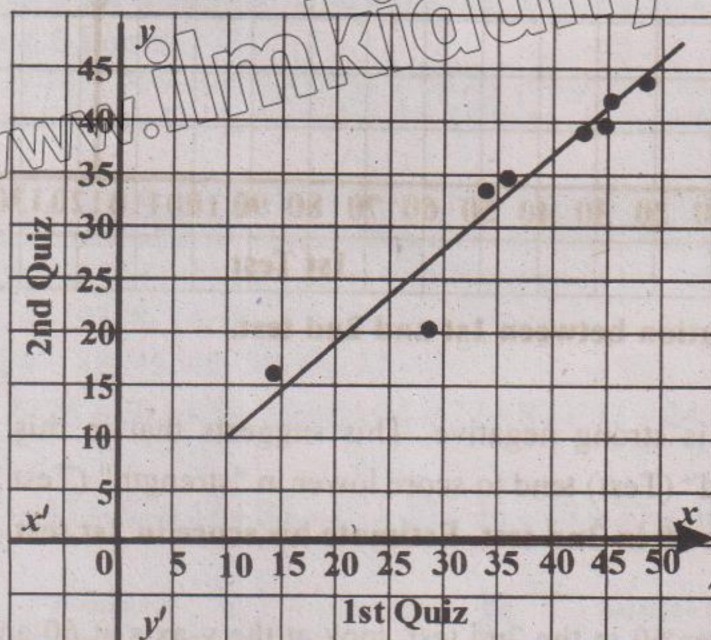
Box-and-Whisker Plot



7. Construct a scatter plot and draw a line of best fit for the following quiz scores for 8 students in a class:

1st quiz	14	28	34	36	43	45	46	49
2nd quiz	16	20	33	35	38	39	42	43

Sol.



Correlation:

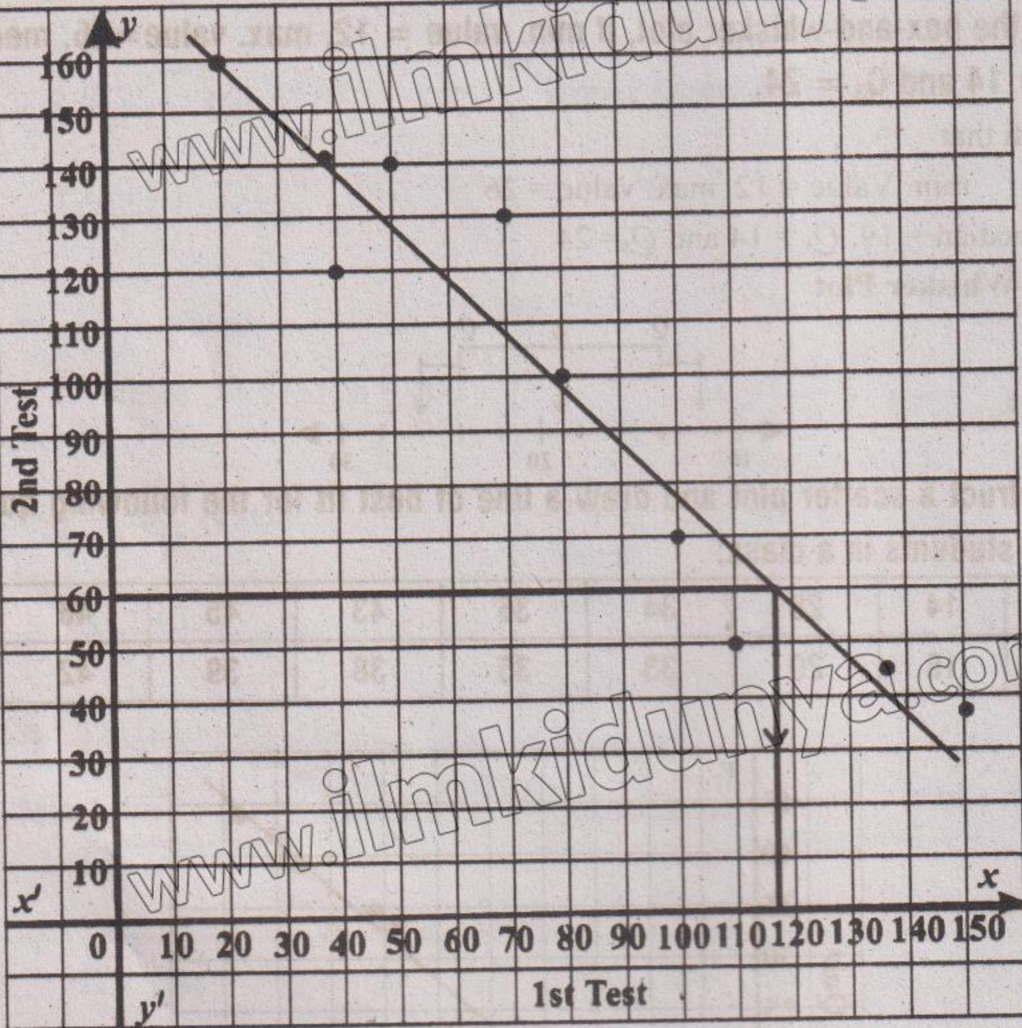
The correlation is strong positive. As the score in the 1st quiz increases, the score in the 2nd quiz also tends to increase.

8. The following table shows the test scores (one measuring speed and the other measuring strength) of football players out of 200:

1st Test	20	38	70	40	100	110	50	136	150	80
2nd Test	160	142	130	120	70	50	140	44	36	100

(i) Construct a scatter plot and draw a line of best fit.

Sol.



(ii) Describe correlation between 1st and 2nd test.

Sol. Correlation:

The correlation is strong negative. This suggests that in this dataset, player who score higher in "speed" (Test) tend to score lower in "strength" (Test 2).

(iii) Abdullah scores 60 in 2nd test. Estimate his score in 1st test.

Sol. Estimation:

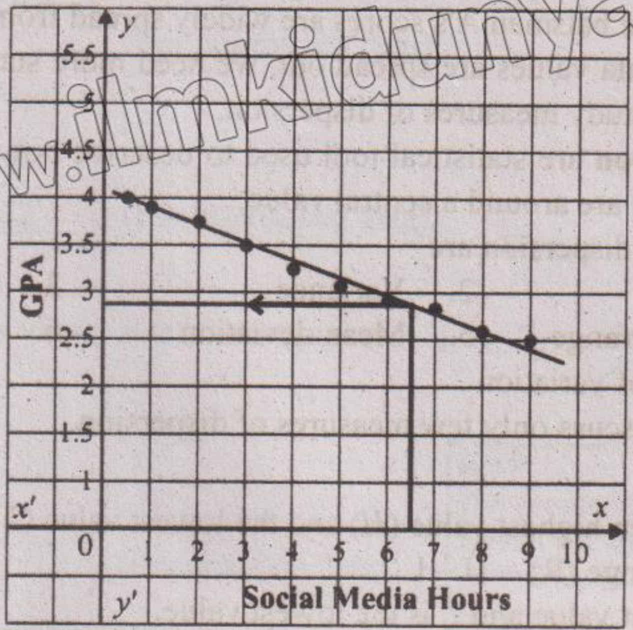
If Abdullah scores 60 in the 2nd test, look at the y-axis at 60 and move horizontally to live of best fit, then down to the x-axis. Based on the trend, his test 1 score would be approximately 110 to 120 ≅ 115.5

9. The following table shows the number of daily social media hours and GPA of university students.

Social Media Hours	0.5	1	2	3	4	5	6	7	8	9
GPA	4.0	3.9	3.7	3.5	3.3	3.1	2.9	2.8	2.6	2.5

(i) Construct a scatter diagram and draw a line of best fit.

Sol.



(ii) Identify the nature of correlation between social media use and GPA.
Sol. Nature of correlation:

This is a negative linear correlation. As social media usage increase, the GPA consistently decreases.

(iii) Estimate the GPA for a student who spends 6.5 hours on social media.
Sol. Estimation for 6.5 hours:

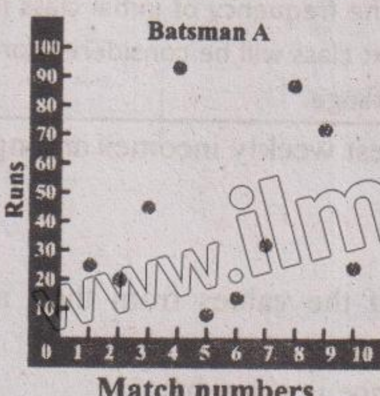
Since the data is very consistent (GPA) drop by roughly 0.1 or 0.2 for every hours. We can look between 6 hours (2.9), the value for 65 hours is exactly in middle is $2.85 \cong 2.9$

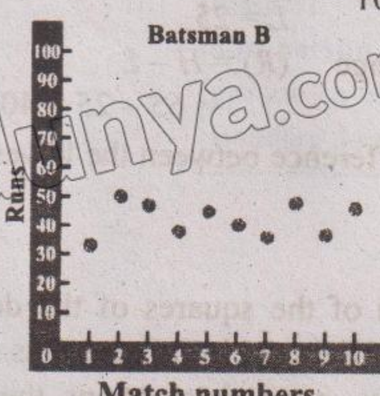
11.4 Measures of Dispersion

The following data shows the runs scored by two batsmen in the last 10 matches:
Batsman *A*: 25, 20, 45, 93, 8, 14, 32, 87, 72, 24
Batsman *B*: 33, 50, 47, 38, 45, 40, 36, 48, 37, 46

$$\text{Mean of Batsman } A = \frac{25+20+45+93+8+14+32+87+72+24}{10} = \frac{420}{10} = 42$$

$$\text{Mean of Batsman } B = \frac{33+50+47+38+45+40+36+48+37+46}{10} = \frac{420}{10} = 42$$





Although both batsmen have the same mean score, batsman B's runs are closely clustered around the mean, while batsman A's scores are widely spread from 0 to 100. Therefore, to understand how the data values are spread out, we need more statistical information. For this purpose, we will study measures of dispersion.

Measures of dispersion are statistical tool used to describe how spread out or scattered the values in a data set are around a central value.

Different measures of dispersion are

1. Range
2. Variance
3. Standard deviation
4. Interquartile range
5. Mean deviation
6. Quartile deviation
7. Coefficient of variation

In this unit, we will discuss only few measures of dispersion.

Range

The difference between highest value (H) and the lowest value (L) is called range.

$$\text{Range (R)} = H - L$$

Where H is the highest value and L is the lowest value.

Example 7: Find the range of the following data:

15, 22, 18, 30, 17, 25

Solution Here $H = 30; L = 15$

Range $(R) = H - L$

$= 30 - 15 = 15$

Skilled practice

What is the range of first 20 prime numbers?

Sol: First 20 prime numbers are

2, 3, 5, 7, 11, 13, ..., 61, 67, 71

Here

Highest Value (H) = 71

Lowest Value (L) = 2

Range (R) = H − L

= 71 − 2

Range (R) = 69

Example 8 The table below shows the weekly incomes (in thousands of rupees) of group of employees in a company: Find the range of the weekly income.

Weekly Income (Rs 1000)	Number of Employees
25 − 35	4
35 − 45	6
45 − 55	5
55 − 65	3

Solution Here $H = 65$

$L = 25$

Range $(R) = H - L$

$= 65 - 25 = 40$

REMEMBER

If the frequency of initial class is zero then the next class will be considered for the calculation of range.

Thus, the difference between the lowest and highest weekly incomes among employees is Rs. 40,000.

Variance

The average of the squares of the deviations of the values from their mean is called variance. It is denoted by σ^2 (read as sigma square).

If $x_1, x_2, \ldots, x_n$, are n observations, then the variance is given by

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} \quad \text{(For ungrouped data)}$$

$$\text{or} \quad \sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2 \quad \text{(For ungrouped data)}$$

Let $x_1, x_2, \ldots, x_m$, be the n values with the frequencies $f_1, f_2, f_3, \ldots, f_n$ respectively.

Then,

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \overline{x})^2}{\sum_{i=1}^{n} f_i} \quad \text{(For grouped data)}$$

Note
The greater the variance, the more spread out the data; the smaller the variance, the more consistent or tightly clustered the values are around the mean..

$$\text{or} \quad \sigma^2 = \frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}\right)^2 \quad \text{(For grouped data)}$$

Standard Deviation

Standard deviation is the positive square root of the average of the squares of deviations of the given values from their mean. It is denoted by σ.

Standard deviation helps us understand the extent to which values differ from the mean.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \quad \text{(For ungrouped data)}$$

$$\text{or } \sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2} \quad \text{(For grouped data)}$$

History!
Karl person was the firs person to use the word standard deviation.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \overline{x})^2}{\sum_{i=1}^{n} f_i}} \quad \text{(For grouped data)}$$

$$\text{or } \sigma = \frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i} - \left[\frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}\right]^2 \quad \text{(For grouped data)}$$

> **Challenge!**
> Can variance be negative?
> Ans: No, variance cannot be negative, because its evaluation involves squares of values and summing of zero or positive values.

Example 9: A factory produces rods with lengths (in cm):

50.1, 49.9, 50.0, 50.2, 49.8

Check the consistency using standard deviation.

Solution $\quad \overline{x} = \dfrac{50.1+49.9+50.0+50.2+49.8}{5} = \dfrac{250}{5} = 50$

x	$x-\overline{x}$	$(x-\overline{x})^2$
50.1	0.1	0.01
49.9	–0.1	0.01
50.0	0	0.00
50.2	0.2	0.04
49.8	–0.2	0.04
		$\sum(x-\overline{x})^2 = 0.10$

$$\sigma^2 = \frac{\sum(x-\overline{x})^2}{n} = \frac{0.10}{5} = 0.02 \Rightarrow \sigma = \sqrt{0.02} = 0.141\text{cm}$$

As the standard deviation is very low, this shows high consistency in rod lengths.

Example 10: The number of carpets sold in each day of a week is 13, 8, 4, 9, 7, 12, 10. Find its variance.

Solution:

x	13	8	4	9	7	12	10	$\Sigma x = 63$
x^2	169	64	16	81	49	144	100	$\Sigma x^2 = 623$

$$\sigma = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

$$= \frac{623}{7} - \left(\frac{63}{7}\right)^2 = 89 - 81 = 8$$

> **Challenge!**
> Can the standard deviation be more than the variance?
> Ans: Yes, it can be more than variance when the variance is between 0 and 1.
> i.e $0 < \sigma^2 < 1$
> Because square root of any decimal number between 0 and 1 is always larger.

Example 11: A financial analyst recorded the closing share prices of several companies on a specific day and grouped them as follows:

Share prices (Rs.)	40–50	50–60	60–70	70–80	80–90	90–100
Frequency	2	3	6	4	2	1

Find the mean, variance and standard deviation. What can you conclude about the volatility in share prices?

Solution:

Price Range (Rs.)	f	x	fx	fx^2
40–50	2	45	90	4050
50–60	3	55	165	9075
60–70	6	65	390	25350
70–80	4	75	300	22500
80–90	2	85	170	14450
90–100	1	95	95	9025
	$\Sigma f = 18$		$\Sigma fx = 1210$	$\Sigma fx^2 = 84450$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{1210}{18} = 67.22$$

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2 = \frac{84450}{18} - (67.22)^2$$

$$= 4691.67 - 4518.52 = 173.15$$

$$\sigma = \sqrt{173.15} = 13.16$$

Standard deviation of Rs. 13.16 shows moderate price volatility.

Example 12: Two students, Ali and Rooma, appeared in 5 Mathematics tests. Their score (out of 100) were as follows:

Ali: 72, 68, 74, 70, 76

Rooma: 72, 65, 80, 68, 75

(i) Who performed better on average?

(ii) Who was more consistent?

Remember!
When comparing two data sets:
- The mean tells us about the average performance.
- The standard deviation tells us about the consistency.

Solution: $\bar{x}(\text{Ali}) = \frac{\sum x}{n} = \frac{72+68+74+70+76}{5} = \frac{360}{5} = 72$

$$\sigma^2 = (\text{Ali}) = \frac{\sum(x-x)^2}{n}$$

$$= \frac{(72-72)^2+(68-72)^2+(74-72)^2+(70-72)^2+(76-72)^2}{5}$$

$$= \frac{(0)^2+(-4)^2+(2)^2+(-2)^2+(4)^2}{5}$$

$$= \frac{0+16+4+4+16}{5}$$

$$= \frac{40}{5} = 8$$

$$\sigma = (\text{Ali}) = \sqrt{8} = 2.83$$

$$\bar{x}(\text{Rooma}) = \frac{\sum x}{n} = \frac{72+65+80+68+75}{5}$$

$$= \frac{360}{5} = 72$$

$$\bar{x}(\text{Rooma}) = \frac{\sum(x-\bar{x})^2}{n}$$

$$= \frac{(72-72)^2+(65-72)^2+(80-72)^2+(68-72)^2+(75-72)^2}{5}$$

$$= \frac{(0)^2+(-7)^2+(8)^2+(-4)^2+(3)^2}{5}$$

$$= \frac{0+49+64+16+9}{5} = \frac{138}{5} = 27.6$$

$$\bar{x}(\text{Rooma}) = \sqrt{27.6} = 5.25$$

Since $\bar{x}$(Ali) = 72, $\bar{x}$ (Rooma) = 72. Both performed equally on average.

σ = (Ali) = 2.83, σ (Rooma) = 5.25

Hence, Ali is more consistent.

Solved Exercise 11.2

1. Find the range of the following data sets:

(i) 63, 89, 98, 125, 79, 108, 117, 60

Sol. Given data set

63, 89, 98, 125, 79, 108, 117, 60

Here

Highest value (H) = 125

Lowest value (L) = 60

Range (R) = H – L

= 125 – 60

Range (R) = 65 Ans.

(ii) 43, 5, 13, 6, 18, 9, 38, 4, 61, 4, 64, 4

Sol: Given data set

43.5, 13.6, 18.9, 38.4, 61.4, 29.4

Here

Highest value (H) = 61.4

Lowest value (L) = 13.6

Range (R) = H – L

= 61.4 – 13.6

Range (R) = 47.8 Ans.

2. If the range and the lowest value of a set of data are 46.7 and 13.4 respectively, then find the highest value.

Sol. Given dat aset

Range (R) = 46.7

Lowest value (L) = 13.4

Range (R) = H – L

46.7 = H – 13.4

46.7+13.4 = H

H = 60.1 Ans.

3. Calculate the range of the following data.

Income (in Rs.)	4000 - 4500	4500 - 5000	5000- 5500	5500- 6000	6000- 6500
No. of warkers	8	12	30	21	6

Sol: Highest Income (H) = 6500

Lowest Income (L) = 4000

$\because$ Range (R) = H – L

= 6500–4000

$\Rightarrow$ Range (R) = 2500 Ans.

4. A group of 7 workers reported the number of items they assembled in a day as: 52, 55, 50, 53, 54, 56, 52

Sol: $\bar{x} = \dfrac{52+55+50+53+54+56+52}{7} = \dfrac{372}{7} = 53.14$

x	$x-\bar{x}$	$(x-\bar{x})^2$
52	52–53.14 = –1.14	$(-1.14)^2 = 1.30$
55	55–53.14 = 1.86	$(1.86)^2 = 3.46$
50	50–53.14 = –3.14	$(-3.14)^2 = 9.86$
53	53–53.14 = –0.14	$(-0.14)^2 = 0.02$
54	54–53.14 = 0.86	$(0.86)^2 = 0.74$
56	56–53.14 = 2.86	$(2.86)^2 = 8.18$
52	52–53.14 = –1.14	$(-1.14)^2 = 1.30$
		$\Sigma(x-\bar{x})^2 = 24.86$

$\because$ Variance $(\sigma^2) = \dfrac{\sum(x-\bar{x})}{n} = \dfrac{24.86}{7}$

$\boxed{\sigma^2 = 3.55}$

$\because$ Standard Deviation $(\sigma) = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n}}$

$= \sqrt{\dfrac{24.86}{7}}$

$= \sqrt{3.55}$

$\boxed{\sigma = 1.88}$

5. A librarian recorded the number of visitors during 5 days of a week. 120, 135, 130, 125, 140

Sol. $\bar{x} = \dfrac{120+135+130++125+140}{5}$

$\bar{x} = \frac{650}{5}$

$\boxed{\bar{x} = 130}$

x	$(x-\bar{x})$	$(x-\bar{x})^2$
120	120–130 = –10	$(-10)^2 = 100$
135	135–130 = 5	$5^2 = 25$
130	130–130 = 0	$0^2 = 0$
125	125–130 = –5	$(-5)^2 = 25$
140	140–130 = 10	$10^2 = 100$
		$\Sigma (x-x)^2 = 250$

$$\text{Variance } (\sigma^2) = \frac{\sum(x-\bar{x})^2}{n}$$

$$= \frac{250}{5}$$

$$\boxed{\sigma^2 = 50}$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

$$= \sqrt{\frac{250}{5}}$$

$$= \sqrt{50}$$

$$\boxed{\sigma = 7.07}$$

6. Find the range, variance and standard deviation of first 23 odd numbers.

Sol. First 23 odd numbers are

1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45

Range

$\because$ Range (R) = Highest value (H) – Lowest value (L)

= 45 – 1

Range (R) = 44

n = 23

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

$$\bar{x} = \frac{1+3+5+\dots+45}{23}$$

$$\bar{x} = \frac{529}{23}$$

$$\boxed{\bar{x} = 23}$$

x	$x-\bar{x}$	$(x-\bar{x})^2$
1	1 − 23 = − 22	$(-22)^2 = 484$
3	3 − 23 = − 20	$(-20)^2 = 400$
5	5 − 23 = − 18	$(-18)^2 = 324$
7	7 − 23 = − 16	$(-16)^2 = 256$
9	9 − 23 = − 14	$(-14)^2 = 196$
11	11 − 23 = − 12	$(-12)^2 = 144$
13	13 − 23 = − 10	$(-10)^2 = 100$
15	15 − 23 = − 8	$(-8)^2 = 64$
17	17 − 23 = − 6	$(-6)^2 = 36$
19	19 − 23 = − 4	$(-4)^2 = 16$
21	21 − 23 = − 2	$(-2)^2 = 4$
23	23 − 23 = 0	$0^2 = 0$
25	25 − 23 = 2	$2^2 = 4$
27	27 − 23 = 4	$4^2 = 16$
29	29 − 23 = 6	$6^2 = 36$
31	31 − 23 = 8	$8^2 = 64$
33	33 − 23 = 10	$10^2 = 100$
35	35 − 23 = 12	$12^2 = 144$
37	37 − 23 = 14	$14^2 = 196$
39	39 − 23 = 16	$16^2 = 256$
41	41 − 23 = 18	$18^2 = 324$
43	43 − 23 = 20	$20^2 = 400$
45	45 − 23 = 22	$22^2 = 484$
		$\Sigma (x-x)^2 = 4048$

Variance $(\sigma^2) = \dfrac{\sum(x-\bar{x})^2}{n}$

$= \dfrac{4048}{23}$

$\boxed{\sigma^2 = 176}$ Ans.

S.D$(\sigma) = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n}}$

$= \sqrt{\dfrac{4048}{23}}$

$= \sqrt{176}$

$\boxed{\sigma = 13.27}$ Ans.

7. **The rainfall recorded in various places of five districts in a week is given below. Find its variance and standard deviation.**

Rainfall (in mm)	42	51	54	61	63	71
Number of places	5	13	4	9	5	4

Sol:

Rain fall (in mm) x	Number of places f	$f.x$	x^2	$f.x^2$
42	5	210	1764	8820
51	13	663	2601	33813
54	4	216	2916	11664
61	9	549	3721	33489
63	5	315	3969	19845
71	4	284	5041	20164
	$\sum f = 40$	$\sum fx = 2237$		$\sum fx^2 = 127795$

$$\text{Variance } (\sigma^2) = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

$$= \frac{127795}{40} - \left(\frac{2237}{40}\right)^2$$

$$= 3194.875 - (55.925)^2$$

$$= 3194.875 - 3127.605$$

$$\boxed{\sigma^2 = 67.27\text{mm}}$$

$$\text{S.D}(\sigma) = \sqrt{\text{Variance}} = \sqrt{67.27}$$

$$\boxed{\sigma = 8.2\text{mm}}$$

8. **Machine A Output (units): 98, 100, 102, 101, 99**
   Machine B Output (units): 95, 100, 105, 90, 110
   (i) Which machine has better performance?
   (ii) Which machine is more consistente?

Sol.

(i) Machine A output Units:

98, 100, 102, 101, 99

$$\text{Now Mean } (\bar{x}_A) = \frac{\sum x}{n}$$

$$\text{Mean } (\bar{x}_A) = \frac{98+100+102+101+99}{5} = \frac{500}{5}$$

$$\boxed{\bar{x}_A = 100}$$

And

$$\sigma^2 \text{ (Machine A)} = \frac{\sum(x-\bar{x})^2}{n}$$

$$= \frac{(98-100)^2+(100-100)^2+(102-100)^2+(101-100)^2+(99-100)^2}{5}$$

$$= \frac{(-2)^2+0^2+2^2+1^2+(-1)^2}{5}$$

$$= \frac{4+0+4+1+1}{5}$$

$$= \frac{10}{5}$$

σ^2 (Machine A) = 2

$$\boxed{S.D\ (\sigma_A) = \sqrt{2} \approx 1.41}$$

(ii) Machine B output Units:

95, 100, 105, 90, 110

$$\text{Now Mean } (\bar{x}_B) = \frac{\sum x}{n}$$

$$\text{Mean } (\bar{x}_B) = \frac{95+100+105+90+110}{5}$$

$$= \frac{500}{5} = 100$$

$$\sigma^2 \text{ (Machine B)} = \frac{\sum(x-\bar{x})^2}{n}$$

$$\sigma^2 = \frac{(95-100)^2+(100-100)^2+(105-100)^2+(90-100)^2+(110-100)^2}{5}$$

$$\sigma^2 = \frac{(-5)^2+(0)^2+(5)^2+(-10)^2+(10)^2}{5}$$

$$\sigma^2 = \frac{25+0+25+100+100}{5} = \frac{250}{5} = 50$$

$\sigma^2 = 50$

$\sigma_B = \sqrt{50}$

S.D(σ_B) = $\sqrt{50} \approx 7.07$

(i) We see that means are equal. Both machines have equal performence.

(ii) Machine A is more consistant

9. The monthly sales (rupees in lacs) for two salespersons over 6 months are:
Person A: 5.5, 5.7, 5.4, 5.6, 5.8, 5.6
Person B: 5.5, 6.5, 4.5, 6.0, 5.0, 6.0
Compare their Performance and consistent

Sol.

- **Person A**

5.5, 5.7, 5.4, 5.6, 5.8, 5.6

Now Mean $(P_A) = \bar{x}_A = \dfrac{\sum x}{n} = \dfrac{5.5+5.7+5.4+5.6+5.8+5.6}{6} = \dfrac{33.6}{6}$

$$\boxed{\bar{x}_A = 5.6}$$

And $\sigma^2(P_A) = \dfrac{\sum(x-\bar{x})^2}{n}$

$$\sigma^2(P_A) = \frac{(5.5-5.6)^2+(5.7-5.6)^2+(5.4-5.6)^2+(5.6-5.6)^2+(5.8-5.6)^2+(5.6-5.6)^2}{6}$$

$$= \frac{(-0.1)^2+(0.1)^2+(-0.2)^2+0^2+(0.2)^2+0^2}{6}$$

$$= \frac{0.01+0.01+0.04+0+0.04+0}{6} = \frac{0.1}{6}$$

$$\boxed{\sigma^2(P_A) \approx 0.0167}$$

$\Rightarrow$ S.D $(\sigma_A) = \sqrt{0.0167}$

≈ 0.129

$$\boxed{\text{S.D } (\sigma_A) \approx 0.13}$$

- **Person B**

5.5, 6.5, 4.5, 6.0, 5.0, 6.0

Now Mean $(P_B) = \bar{x}_B = \dfrac{5.5+6.5+4.5+6.0+5.0+6.0}{6} = \dfrac{33.5}{6}$

$$\boxed{\bar{x}_B = 5.58}$$

And.

$\sigma^2(P_B) = \dfrac{\sum(x-\bar{x})^2}{n}$

$$= \frac{(5.5-5.58)^2+(6.5-5.58)^2+(4.5-5.58)^2+(6-5.58)^2+(5-5.58)^2+(6-5.58)^2}{6}$$

$$= \frac{(-0.08)^2+(0.92)^2+(-1.08)^2+(0.42)^2+(-0.58)^2+(0.42)^2}{6}$$

$$= \frac{0.0064 + 0.8464 + 1.1664 + 0.1764 + 0.3364 + 0.1764}{6}$$

$$= \frac{2.7084}{6}$$

$$\sigma^2 (P_B) = 0.4514$$

$$\Rightarrow \boxed{S.D\ (\sigma_B) = \sqrt{0.4514}}$$

$$\boxed{S.D\ (\sigma_B) \approx 0.672}$$

We see that person A has higher mean than person B. So, person A is better in performance.

And

Person A has lower S.D = 0.13. So, person A is more consistent.

10. The table given below shows the daily wages of workers in a textile mill, grouped into six income brackets:

Daily wage (Rs)	800–1000	1000–1200	1200–1400	1400 – 1600	1600 – 1800	1800 – 2000
Frequency	2	4	6	8	2	1

Calculate the mean, variance and standard deviation of the wages.

Sol:

Weekly Wage (Rs)	Frequency (f)	Midpoint (x)	f.x	x^2	$f.x^2$
800 – 1000	2	$\frac{1800}{2} = 900$	1800	810000	1620000
1000 – 1200	4	$\frac{2200}{2} = 1100$	4400	1210000	4840000
1200 – 1400	6	$\frac{2600}{2} = 1300$	7800	1690000	10140000
1400 – 1600	8	$\frac{3000}{2} = 1500$	12000	225000000	18000000
1600 – 1800	2	$\frac{3400}{2} = 1700$	3400	2890000	5780000
1800 – 2000	1	$\frac{3800}{2} = 1900$	1900	3610000	3610000
	$\Sigma f = 23$		$\Sigma f x = 31300$		$\Sigma f x^2 = 43990000$

- **Mean**

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{31300}{23}$$

$$\boxed{\bar{x} = 1360.87}$$

- **Variance**

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

$$= \frac{43990000}{23} - \left(\frac{31300}{23}\right)^2$$

$$= 1912608.70 - (1360.87)^2$$

$$= 1912608.70 - 1851967.16$$

$$\boxed{\sigma^2 = 60642.72}$$

- **Standard Deviation**

$$S.D = \sqrt{\text{Variance}}$$

$$= \sqrt{60642.72}$$

$$\boxed{S.D = 246.26}$$

11. A company forecasts monthly sales (rupees in millions): 15, 18, 14, 20, 13. Find variability in sales predictions.

Sol: Monthly sales (rupees in millions):

15, 18, 14, 20, 13

To check variability in sales predictions we find mean, variance and S.D of dataset.

- **Mean:**

$$\bar{x} = \frac{15+18+14+20+13}{5} \quad \because \text{ no of values} = n = 5$$

$$\bar{x} = \frac{80}{5}$$

$$\boxed{\bar{x} = 16}$$

- **Variance:**

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

$$= \frac{(15-16)^2 + (18-16)^2 + (14-16)^2 + (20-16)^2 + (13-16)^2}{5}$$

$$= \frac{(-1)^2 + 2^2 + (-2)^2 + (4)^2 + (-3)^2}{5}$$

$$= \frac{1+4+4+16+9}{5} = \frac{34}{5}$$

$$\boxed{2 = 6.8}$$

- **Standard Deviation:**

$$S.D = \sqrt{\text{Variance}}$$

$$= \sqrt{6.8}$$

$$\boxed{\sigma = 2.61}$$

Result: With a S.D = 2.61 relative to mean= 16.
The sales predictions show moderate variability.

12. Unemployment rates (%) in five provinces is 5.2, 6.0, 4.8, 5.5, 6.2. Calculate standard deviation and describe is there balanced unemployment rate?

Sol. The unemployment rates (%) are:

5.2, 6.0, 4.8, 5.5, 6.2

We find here mean, variance and S.D.

- **Mean:**

$$\bar{x}=\frac{\sum x}{n}=\frac{5.2+6.0+4.8+5.5+6.2}{5} \qquad \because \text{no of values} = n = 5$$

$$\bar{x}=\frac{27.7}{5}$$

$\boxed{\bar{x}=5.54}$ (in %)

- **Variance:**

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

$$=\frac{(5.2-5.54)^2+(6.0-5.54)^2+(4.8-5.54)^2+(5.5-5.54)^2+(6.2-5.54)^2}{5}$$

$$=\frac{(-0.34)^2+(0.46)^2+(-0.74)^2+(-0.04)^2+(0.66)^2}{5}$$

$$=\frac{0.1156+0.2116+0.5476+0.0016+0.4356}{5}=\frac{1.312}{5}$$

$\boxed{\sigma^2=0.2624}$ (in %)

- **S.D:**

$$\text{S.D} = \sqrt{\text{Variance}}$$

$$\sigma=\sqrt{0.2624}$$

$\boxed{\sigma=0.51\%}$

Result: The unemployment rate is relatively fairly balanced because there is low dispersion between the provinces.

13. Find variance and standard deviation:

(i) $\Sigma x = 45, \Sigma x^2 = 421, n = 5$

Sol. Given that

$\Sigma x = 45, \Sigma x^2 = 421, n = 5$

$$\text{variance } (\sigma^2) = \frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2$$

$$= \frac{421}{5} - \left(\frac{45}{5}\right)^2$$
$$= 84.2 - 9^2$$
$$= 84.2 - 81$$
$$\boxed{\sigma^2 = 3.2}$$

$$\text{S.D } (\sigma) = \sqrt{\text{Variance}}$$
$$= \sqrt{3.2}$$
$$\boxed{\sigma = 1.79}$$

(ii) $\Sigma f x = 210, \Sigma f x^2 = 7560, \Sigma f = 6$

Sol. Given that

$\sum fx = 210, \sum fx^2 = 7560, \sum f = 6$

$$\text{variance } (\sigma^2) = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$
$$= \frac{7560}{6} - \left(\frac{210}{6}\right)^2$$
$$\sigma^2 = 1260 - 35^2$$
$$\sigma^2 = 1260 - 1225$$
$$\boxed{\sigma^2 = 35}$$
$$\because \text{S.D } (\sigma) = \sqrt{\text{Variance}}$$
$$= \sqrt{35}$$
$$\boxed{\sigma = 5.92}$$

(iii) $\overline{x} = 18, \sum fx^2 = 1670, \sum f = 5$

Sol. Given that

$\overline{x} = 18, \sum fx^2 = 1670, \sum f = 5$

$$\because \quad \text{variance } (\sigma^2) = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$
$$= \frac{1670}{5} - (\overline{x})^2$$
$$= \frac{1670}{5} - (18)^2$$
$$= 334 - 324$$
$$\boxed{\sigma^2 = 10}$$
$$\because \text{S.D } (\sigma) = \sqrt{\text{Variance}}$$
$$= \sqrt{10}$$
$$\boxed{\sigma = 3.16}$$

Solved Review Exercise 11

1. Four possible answers are given for the following questions. Choose the correct answer.

(i) ______ is used to get the cumulative frequencies.

(a) Addition (b) square root (c) multiplication (d) division

(ii) Second quartile represents:

(a) mean (b) mode (c) median (d) variance

(iii) First quartile divides the data into ______ equal parts.

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

(iv) Difference between the highest and the lowest values is called:

(a) mean (b) variance (c) range (d) standard deviation

(v) Scatter diagram represents a relationship between _____ variables.

(a) five (b) two (c) three (d) four

(vi) _______ is measure of dispersion:

(a) mean (b) median (c) mode (d) variance

(vii) **Positive square root of variance is called:**

(a) mean (b) median (c) standard deviation (d) range

(viii) _______ **is not measure of dispersion.**

(a) range (b) arithmetic mean (c) variance (d) standard deviation

(ix) **Variance of the data 8, 8, 8, 8, 8, 8 is:**

(a) 0 (b) 16 (c) 8 (d) 48

(x) **Range of first 20 natural numbers is:**

(a) 20 (b) 10 (c) 19 (d) 30

ANSWERS:

(i) Addition (ii) Median (iii) Four (iv) Range (v) Two

(vi) Variance (vii) Standard Deviation (viii) Arithmatic Mean (ix) 0 (x) 19

2. The following results for the long jump were recorded:

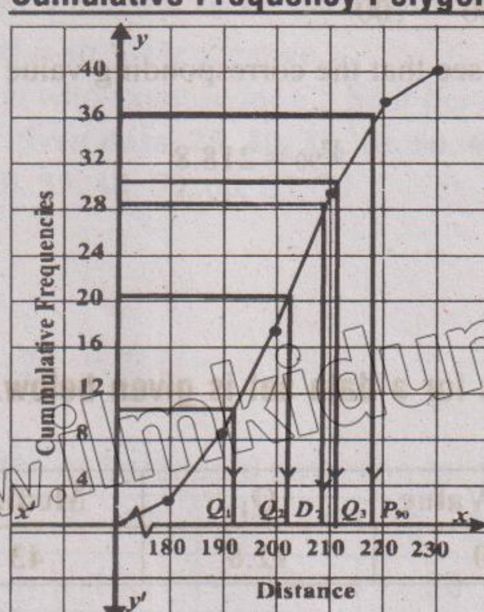
Distance (in cm)	170 - 180	180-190	190-200	200-210	210-220	220-230
f	2	6	9	12	8	3

Construct the cumulative frequency polygon and interpret median, Q_1, Q_3, D_7, P_{90} and interquartile range.

Sol.

Distance (in cm)	*f*	*cf*
170 – 180	2	2
180 – 190	6	2 + 6 = 8
190 – 200	9	8 + 9 = 17
200 – 210	12	17 + 12 = 29
210 – 220	8	29 + 8 = 37
220 – 230	3	37 + 3 = 40
	$\sum f = 40$	

Cumulative Frequency Polygon



Median The position of median is $\frac{n}{2}=\frac{40}{2}=20$ (20^{th} value). Draw a horizontal line from 20 on the y-axis until it meets the curve and then draw vertical line from this point to the x-axis.

Now take the reading on the x-axis, where the vertical line hits the x-axis. So,

$$\text{Median} = 202.5$$

Lower Quartile (Q_1)

The position of Q_1 is $\frac{n}{4}=\frac{40}{4}=10$ (10^{th} value).

Now from the graph we see that the corresponding value of 10 on y-axis is 192.2 on x-axis.

$$Q_1 = 192.2$$

Upper Quartile (Q_3)

The position of Q_3 is $\frac{3n}{4}=\frac{3(40)}{4}=30$ (10^{th} value).

Now from the graph we see that the corresponding value of 30 on y-axis is 211.3 on x-axis.

$$Q_3 = 211.3$$

7^{th} Decile (D_7)

The position of D_7 is $\frac{7n}{10}=\frac{7(40)}{10}=28$ (28^{th} value).

Now from the graph we see that the corresponding value of 28 on y-axis is 209.2 on x-axis.

$$D_7 = 209.2$$

90^{th} Percentile (P_{90})

The position of P_{90} is $\frac{90n}{100}=\frac{90(40)}{100}=36$ (36^{th} value)

Now from the graph we see that the corresponding value of 36 on y-axis is 218.8 on x-axis.

$$P_{90} = 218.8$$

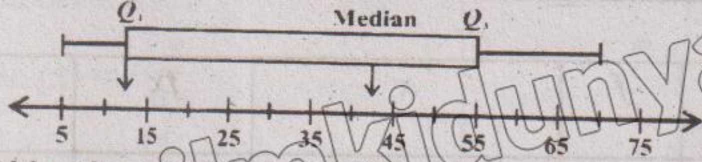
Interquartile Range (IQR)

$$\because \quad IQR = Q_3 - Q_1$$
$$= 211.3 - 192.2$$
$$IQR = 19.1$$

3. The summary statistics for a data set is given below. Show it with a box-and-whisker plot.

Min. Value	Max. Value	Q_1	Median	Q_3
5	70	12.6	43	55.6

Solution:



4. **The following table shows the weekly hours spent exercising and resting heart rate (beats per minute) for ten individuals:**

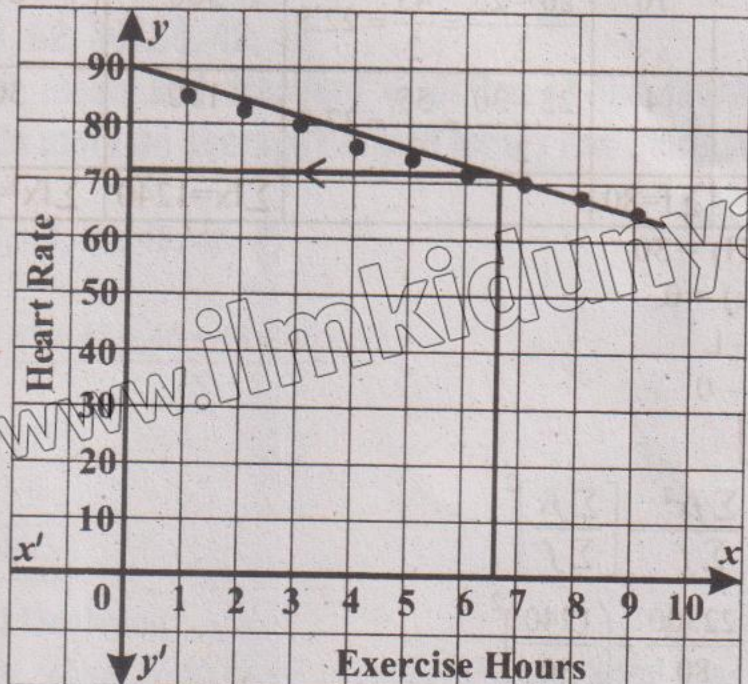
Excersice Hours	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Heart Rate	**90**	**85**	**83**	**80**	**76**	**74**	**72**	**70**	**68**	**65**

(i) Plot the data on a scatter diagram and draw a line of best fit.

(ii) State the type of correlation observed.

(iii) Predict the heart rate of Sakeena who exercises for 6.5 hours per week.

Solution: (i)



(ii) The correlation observed is strongly negative.

(iii) The heart rate of Sakeena who exercise for 6.5 hour per week is 71.

5. **Find the range for the given data, 25, 30, 35, 40, 50, 60, 65, 75**

Solutions: Given data, 25, 30, 35, 40, 50, 60, 65, 75

Highest value (H) = 75

Lowest value (L) = 25

Range (R) = H – L

R = 75 – 25

$\boxed{R = 50}$

6. **Calculate range, variance and standard deviation for the following data set:**

Class Interval	**0 – 5**	**5 – 10**	**10 – 15**	**15 – 20**	**20 – 25**	**25 – 30**
f	**4**	**12**	**20**	**24**	**16**	**4**

Sol:

Class interval	f	x	fx	fx^2
0 – 5	4	$\frac{0+5}{2}=\frac{5}{2}=2.5$	10	25
5 – 10	12	$\frac{5+10}{2}=\frac{15}{2}=7.5$	90	675
10 – 15	20	$\frac{10+15}{2}=\frac{25}{2}=12.5$	250	3125
15 – 20	24	$\frac{15+20}{2}=\frac{35}{2}=17.5$	420	7350
20 – 25	16	$\frac{20+25}{2}=\frac{45}{2}=22.5$	360	8100
25 – 30	4	$\frac{25+30}{2}=\frac{55}{2}=27.5$	110	3025
	$\sum f=80$		$\sum fx=1240$	$\sum fx^2=22300$

Highest Value (H) = 30

Lowest Value (L) = 0

- Range (R) = H – L

  R = 30 – 0

  $\boxed{R = 30}$

- Variance $(\sigma^2) = \frac{\sum fx^2}{\sum f} - \left[\frac{\sum fx}{\sum f}\right]^2$

  $= \frac{22300}{80} - \left(\frac{1240}{80}\right)^2$

  $= 278.75 - (15.5)^2$

  $= 278.75 - 240.25$

  Variance $(\sigma^2) = 38.5$

- Standard Deviation $(\sigma) = \sqrt{\text{Variance}}$

  $= \sqrt{38.5}$

  $\because$ S.D $(\sigma) = 6.20$

7. The sum of 5 numbers is 45 and the sum of their squares is 421. Find the mean and standard deviation of the data.

Sol: Let 'x' represent numbers.

then

Sum of numbers = $\sum x = 45$

Sum of their squares = $\sum x^2 = 421$

$\because$ Total numbers = n = 5

$$\because \text{Mean } (\bar{x}) = \frac{\sum x}{n}$$

$$\Rightarrow \quad \bar{x} = \frac{45}{n}$$

$$\Rightarrow \quad \boxed{\bar{x} = 9}$$

$$\because \text{S.D } (\sigma) = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \sqrt{\frac{421}{5} - \left(\frac{45}{5}\right)^2}$$

$$= \sqrt{84.2 - 9^2}$$

$$= \sqrt{84.2 - 81}$$

$$= \sqrt{3.2}$$

$$\boxed{\text{S.D } (\sigma) = 1.7889}$$

8. The monthly household expenses (rupees in thousands) of two families for 6 months were:
Family A: 45, 47, 46, 48, 46, 47
Family B: 38, 52, 40, 50, 42, 49
Calculate the mean and standard deviation of the monthly expenses. Which family spends more on average? Which family has more stable expenses?

Sol:

- Family A: 45, 47, 46, 48, 46, 47

Now

$$\text{Mean } (\bar{x}_A) = \frac{45 + 47 + 46 + 48 + 46 + 47}{6}$$

$$= \frac{279}{6}$$

$$\boxed{\bar{x}_A = 46.5}$$

And Standard Daviation

$$\sigma^2{}_A = \frac{(45-46.5)^2 + (47-46.5)^2 + (46-46.5)^2 + (48-46.5)^2 + (46-46.5)^2 + (47-46.5)^2}{6}$$

$$= \frac{(-1.5)^2 + (0.5)^2 + (-0.5)^2 + (1.5)^2 + (-0.5)^2 + (0.5)^2}{6}$$

$$\sigma^2{}_A = \frac{2.25 + 0.25 + 0.25 + 2.25 + 0.25 + 0.25}{6}$$

$$= \frac{5.5}{6}$$

$$\sigma^2{}_A = 0.9167$$

$$\Rightarrow \quad \text{S.D } \sigma_A = \sqrt{0.9167}$$

$$\boxed{\text{S.D } (\sigma_A) = 0.96}$$

- Family B: 38, 52, 40, 50, 42, 49

Now

$$\text{Mean } (\bar{x}_B) = \frac{38+52+40+50+42+49}{6}$$

$$= \frac{271}{6}$$

$$\boxed{\bar{x}_B = 45.17}$$

And Standard Daviation

$$\sigma^2_B = \frac{(38-45.17)^2+(52-45.17)^2+(40-45.17)^2+(50-45.17)^2+(42-45.17)^2+(49-45.17)^2}{6}$$

$$= \frac{(-7.17)^2+(6.83)^2+(-5.17)^2+(4.83)^2+(-3.17)^2+(3.83)^2}{6}$$

$$= \frac{51.41+46.65+26.73+23.33+10.05+14.67}{6}$$

$$= \frac{271}{6}$$

$$\sigma^2_B = 28.81$$

$$\Rightarrow \text{S.D } (\sigma_B) = \sqrt{28.81}$$

$$\boxed{\text{S.D } (\sigma_B) = 5.37}$$

Result Family A spends more on averrage, but family B is more stable.

9. The daily wages of 40 workers in a factory are grouped as follows:

Daily Wages (Rs.)	1000–1200	1200–1400	1400–1600	1600–1800	1800–2000
No. of Workers	6	10	12	8	4

Find the mean, variance and standard deviation of the daily wages.

Sol.

Daily Wages	f	x	fx	x^2	fx^2
1000 – 1200	6	1100	6600	1210000	7260000
1200 – 1400	10	1300	13000	1690000	16900000
1400 – 1600	12	1500	18000	2250000	27000000
1600 – 1800	8	1700	13600	2890000	23120000
1800 – 2000	4	1900	7600	3610000	14440000
	$\Sigma f = 40$		$\Sigma fx = 58800$		$\Sigma fx^2 = 88720000$

- Mean $(\overline{x}) = \dfrac{\sum x}{\sum f} = \dfrac{58800}{40} = 1470$

- Variance $(\sigma^2) = \dfrac{\sum fx^2}{\sum f} - \left(\dfrac{\sum fx}{\sum f}\right)^2$

$$= \frac{88720000}{40} - \left(\frac{58800}{40}\right)^2$$

$$= 2218000 - 2160900$$

Variance $(\sigma^2) = 57100$

- S.D $(\sigma) = \sqrt{57100}$

S.D $(\sigma) = 238.96$

Objective Type Questions

Additional MCQ's

Multiple Choice Questions prepared in the light of new Examination Techniques
(Knowledge, Understanding, Application, Analytical & Conceptual)

11.1 Construction of Cumulative Frequencies

Encircle the correct option.

1. **A table of cumulative frequenceis, also known as:**
   (a) Ogive (b) Cumulative Frequency Polygon
   (c) Cumulative Frequency Distribution (d) Frequency Distribution
2. **A cumulative frequency polygon is also known as:**
   (a) Ogive (b) Table of Cumulative Frequencies
   (c) Cumulative Frequency Distribution (d) Frequency Distribution
3. **Number of methods for construction of an ogive:**
   (a) 1 (b) 2 (c) 3 (d) 4
4. **An ogive helps to estimate:**
   (a) Medians (b) Quartiles (c) Deciles (d) All of these
5. **The middle most value in a dataset arranged in ascending or descending order is known as:**
   (a) Mode (b) Median (c) Quartile (d) Decile
6. **The quartile named as lower quartile is:**
   (a) Q_1 (b) Q_2 (c) Q_3 (d) Q_4
7. **The quartile named as upper quartile is:**
   (a) Q_1 (b) Q_2 (c) Q_3 (d) Q_4
8. **The median divides an arranged data into ______ equal parts:**
   (a) 1 (b) 2 (c) 3 (d) 4

9. **Deciles divide an arranged data into ______ equal parts:**
(a) 2 (b) 4 (c) 10 (d) 100

10. **Percentiles divide an arranged data into ______ equal parts:**
(a) 2 (b) 4 (c) 10 (d) 100

11. **IQR = ______ ?**
(a) $Q_3 - Q_1$ (b) $Q_2 - Q_1$ (c) $Q_1 - Q_3$ (d) $Q_1 - Q_2$

12. **The distance between the first quartile Q_1 and the 3rd quartile Q_3 is known as:**
(a) IQR (b) Median (c) Mode (d) Mean

11.2 Box-and-Whiker Plot

13. **The box in Box-and-Whisker Plot represents:**
(a) Q_1 (b) Q_2 (c) Q_3 (d) IQR

14. **A line within the box represents:**
(a) Q_1 (b) Median (c) Q_3 (d) IQR

15. **Lines extending from the box to minimum and maximum values are called:**
(a) Box (b) Median (c) Whiskers (d) IQR

16. **The median of the dataset 2, 3, 4, 5, 6 is:**
(a) 2 (b) 3 (c) 4 (d) 5

17. **Median = $\left(\frac{n+1}{2}\right)^{th}$ value, if n is:**
(a) Even (b) Odd (c) Prime (d) Rational

11.3 Scatter Diagram

18. **A scatter diagram is a graphical tool used to display the relationship between ______ variables:**
(a) 1 (b) 2 (c) 3 (d) 4

19. **No of categories of correlation is:**
(a) 1 (b) 2 (c) 3 (d) 4

20. **If there is not a consistent patteren between two variable then it is called:**
(a) No correlation (b) Positive correlation
(c) Negative correlation (d) none

21. **If one variable increases as the other increases, then it is:**
(a) No correlation (b) Positive correlation
(c) Negative correlation (d) None

22. **If one variable increases as the other decreases, then it is:**
(a) No correlation (b) Positive correlation
(c) Negative correlation (d) None

23. **If $Q_3 = 75$ and $Q_1 = 25$, then IQR = ______ ?**
(a) 75 (b) 25 (c) 100 (d) 50

24. **If IQR = 200, $Q_1 = 25$, then Q_3 = ______ ?**
(a) 25 (b) 200 (c) 225 (d) 250

11.4 Measures of Dispersion

25. ______ is a measure of dispersion from the following:
(a) Range (b) Correlation (c) Mean (d) Median

26. The range of first 10 odd numbers is:
(a) 1 (b) 9 (c) 18 (d) 27

27. The range is the ______ of min. Value of a data set:
(a) Product (b) Sum (c) Difference (d) Mean

28. If Range (R) = 50, Highest value (H) = 80, then lowest value (L) = _____?
(a) 30 (b) 50 (c) 80 (d) 130

29. Range (R) = _____?
(a) $H \times L$ (b) $\frac{H}{L}$ (c) $H + L$ (d) $H - L$

30. The range of a dataset cannot be ______?
(a) Prime number (b) Even Number (c) Negative Number (d) Positive Integer

31. Notation used to represent variance:
(a) σ (b) σ^2 (c) $\overline{x}$ (d) $\tilde{x}$

32. The formula to calculate variance of ungrouped data is:
(a) $\frac{\sum x}{n}$ (b) $\frac{\sum fx}{\sum f}$ (c) $\frac{\sum(x-\overline{x})^2}{n}$ (d) $\frac{\sum(x-\overline{x})}{n}$

33. The formula $\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2$ is used to calculate:
(a) Mean (b) Median (c) Variance (d) S.D

34. If $\sum(x-\overline{x})^2 = 35$ and $n = 7$, then $\sigma^2 = ?$
(a) 7 (b) 35 (c) 42 (d) 5

35. If $\sum x^2 = 100$, $n = 20$ and $\overline{x} = 2$, then $\sigma^2 = ?$
(a) 1 (b) 2 (c) 3 (d) 4

36. If $\sum x = 54$ and $n = 9$ for a dataset, then $\overline{x} = ?$
(a) 6 (b) 45 (c) 63 (d) 75

37. The formula to calculate mean is:
(a) $\frac{\sum(x-\overline{x})^2}{n}$ (b) $\frac{\sum x^2}{n}$ (c) $\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2$ (d) $\frac{\sum x}{n}$

38. The formula to calculate mean for grouped data is:
(a) $\frac{\sum x}{n}$ (b) $\frac{\sum fx}{\sum f}$ (c) $\frac{\sum x^2}{n}$ (d) $\frac{\sum fx^2}{\sum fx}$

39. Mean of a dataset cannot be ______?
(a) Negative (b) Positive (c) Integer (d) None of these

40. Variance of a dataset cannot be ______?
(a) Negative (b) Positive (c) Even (d) Odd

41. **The formula $\dfrac{\sum fx^2}{\sum f}-\left(\dfrac{\sum fx}{\sum f}\right)^2$ is used to calculate:**
(a) Mean (b) Median (c) Variance (d) S.D

42. **The formula used to calculate variance of a grouped data:**
(a) $\dfrac{\sum fx^2}{\sum f}-\left(\dfrac{\sum fx}{\sum n}\right)^2$ (b) $\dfrac{\sum fx}{\sum f}$ (c) $\sqrt{\dfrac{\sum fx}{\sum f}-\left(\dfrac{\sum fx}{\sum n}\right)^2}$ (d) $\dfrac{\sum(x-\overline{x})^2}{n}$

43. **If $\sum fx^2 = 1670$, $\sum fx = 90$ and $\sum f = 5$, then $\sigma^2 =$ _____ ?**
(a) 5 (b) 10 (c) 15 (d) 20

11.4.3 Standard Deviation (S.D)

44. **Notation used to represent S.D is:**
(a) σ (b) σ^2 (c) $\overline{x}$ (d) $\tilde{x}$

45. **S.D is the positive square root of _____?**
(a) Mean (b) Median (c) Variance (d) Quartiles

46. **If variance (σ^2) = 225, then S.D (σ) = _____?**
(a) 5 (b) 10 (c) 15 (d) 20

47. **The formula to calculate S.D of ungrouped data is:**
(a) $\dfrac{\sum(x-\overline{x})^2}{n}$ (b) $\dfrac{\sum fx}{\sum f}$ (c) $\sqrt{\dfrac{\sum(x-\overline{x})^2}{n}}$ (d) $\sqrt{\dfrac{\sum x^2}{n}}$

48. **While comparing two datasets, mean tells us:**
(a) Average Performance (b) Consistency
(c) Middle value (d) Spreadness of datasets

49. **While comparing two datasets, S.D tells us:**
(a) Average Performance (b) Consistency
(c) Middle Value (d) Max Value

50. **Variance of first 10 natural numbers is:**
(a) 5 (b) 7.5 (c) 8.25 (d) 9.5

ANSWERS:

1. Cumulative Frequencies Distribution			2. Ogive	3. 2
4. All of these		5. Median	6. Q_1	7. Q_3
8. 2	9. 10	10. 100	11. $Q_3 - Q_1$	12. IQR
13. IQR	14. Median	15. Whiskers	16. 4	17. Odd
18. 2	19. 3	20. No correlation	21. Positive correlation	
22. Negative correlation		23. 50	24. 225	25. Range
26. 18	27. Difference	28. 30	29. H − L	
30. negative number		31. σ^2	32. $\dfrac{\sum(x-\overline{x})^2}{n}$	33. Variance
34. 5	35. 1	36. 6	37. $\dfrac{\sum x}{n}$	38. $\dfrac{\sum fx}{\sum f}$

39. None of these 40. Negative 41. Variance 42. $\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$

43. 10 44. σ 45. Variance 46. 15 47. $\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$

48. Average Performance 49. Consistency 50. 8.25

Additional Short Question

Short Answer Questions prepared in the light of new Examination Techniques
(Knowledge, Understanding, Application, Analytical & Conceptual)

11.1 Construction of Cumulative Frequencies

Give short answers.

1. **Define Cumulative Frequency.**

Ans. Cumulative frequency is the running total of frequencies, where each frequency is added to the sum of all frequencies before it.

2. **What is cumulative frequency polygon?**

Ans. A cumulative frequency polygon is a graphical representation that shows the cumulative total of frequencies up to each class boundary in a frequency distribution.

3. **Name the methods for construction of cumulative frequency polygon.**

Ans. There are two methods for construction of cumulative frequency polygon.

(a) Less than method (b) More than method

4. **Define median.**

Ans. The median is the middle value in a set of data when the values are arranged in ascending or descending order.

5. **Find median of the dataset 5, 4, 9, 10, 12, 1, 7.**

Ans. Given data set arranged in ascending order:

1, 4, 5, 7, 9, 10, 12

Here

$n = 7$

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

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

$$= \left(\frac{8}{2}\right)^{th} \text{value}$$

Median = 4^{th} value

$\Rightarrow$ $\boxed{\text{Median} = 7}$

6. Write the formula to find median of ungrouped data, when 'n' is odd.

Ans. Formula

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

7. Find the median of first 11 whole number.

Ans. First 11 whole numbers are:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Here $n = 11$

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

$$= \left(\frac{11+1}{2}\right)^{th} \text{ value}$$

$$= \left(\frac{12}{2}\right)^{th} \text{ value}$$

$$\boxed{\text{Median} = 6^{th} \text{ value} \Rightarrow \text{Median} = 5}$$

8. What is the median of dataset 2.5, 3.5, 4.5, 5.5, 6.5.

Sol. Here

$n = 5$

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

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

$$= \left(\frac{6}{2}\right)^{th} \text{ value} = 3^{rd} \text{ value}$$

$$\Rightarrow \boxed{\text{Median} = 4.5}$$

9. What are quartiles?

Ans. Quartiles are values that divide a dataset into four equal parts each containing 25 % of the data.

10. Write three main quartiles for a dataset.

Ans. Three main quartiels are:

a) First Quartile (Q_1) **b)** Second Quartile (Q_2) **c)** Third Quartile (Q_3)

11. Why we use quartiels?

Sol. Quartiles help us to understand the spread and skewness of data better than a simple average can.

12. Find Q_1 and Q_3 for the dataset 2, 5, 8, 10, 12, 15, 20, 22.

Sol. Given data set in ascending order:

2, 5, 8, 10, 12, 15, 20, 22

Here

$n = 8$

- First Quartile (Q_1)

$$\text{The position of } Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ value}$$

$$Q_1 = \left(\frac{8+1}{4}\right)^{th} \text{ value} = \left(\frac{9}{4}\right)^{th} \text{ value} = 2.25^{th} \text{ value}$$

$$\Rightarrow Q_1 = 2^{nd} \text{ value} + 0.25(3^{rd} \text{ value} - 2^{nd} \text{ value})$$

$$= 5 + 0.25\,(8-5) = 5 + 0.25\,(3)$$

$$= 5 + 0.75$$

$$\Rightarrow \boxed{Q_1 = 5.75}$$

- <u>Third Quartile</u> (Q_3)

The position of $Q_3 = \left(\frac{3(n+1)}{4}\right)$ value

$$= \left(\frac{3(8+1)}{4}\right)^{th} \text{ value} = \left(\frac{27}{4}\right)^{th} \text{ value} = 6.75^{th} \text{ value}$$

$$Q_3 = 6^{th} \text{ value} + 0.75\ (7^{th} \text{ value} - 6^{th} \text{ value})$$
$$= 15 + 0.75\ (20 - 15) = 15 + 0.75\ (5)$$
$$= 15 + 3.75$$
$$\boxed{Q_3 = 18.75}$$

13. What is the value of IQR for the data set 2, 5, 8, 10, 12, 15, 20, 22

Ans. Since for given dataset

$Q_1 = 5.75$ and $Q_3 = 18.75$

$\Rightarrow$ IQR $= Q_3 - Q_1$

$= 18.75 - 5.75$

$\boxed{\text{IQR} = 13}$

11.2 Box-and-Whisker Plot

14. What is Box-and-Whisker Plot?

Ans. A box-and-whisker plot is a graphical representation of a data set five number summary (minimum, first quartile, median, third quartile, and maximum).

15. Why we use Box-and-Whisker plot?

Ans. It is used to visualize data distribution and comparing data sets.

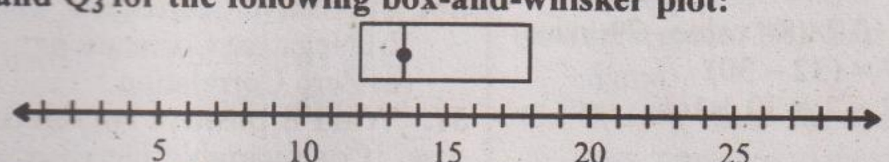
16. What does the box represents in Box-and-Whisker plot?

Ans. It represents the interquartile range (IQR) which is 50% of the data making box edges the first quartile Q_1 and third quartile Q_3.

17. What are whiskers in box-and-whisker plot?

Ans. Lines extending from the box to the minimum and maximum values, showing the overall spread of the data, are whiskers in box-and-whisker plot.

18. Find Q_1 and Q_3 for the following box-and-whisker plot:



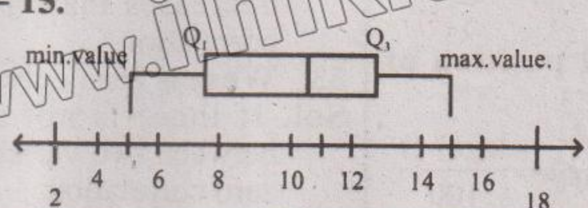
Sol. From the box-and-whisker plot

First Edge of box $= Q_1 = 12$

Second Edge of box $= Q_3 = 18$

19. Plot the box-and-whisker plot, if min. value = 2, max. value = 18, median = 11, $Q_1 = 4.5$ and $Q_3 = 15$.

Sol.



20. Find IQR Value if $Q_1=73$ and $Q_3=80$.

Ans. Given data

$Q_3 = 80,\ Q_1 = 73$

$\Rightarrow\ IQR = Q_3 - Q_1$

$= 80 - 73$

$\boxed{IQR = 7}$

21. What are Deciles?

Sol. The values that divide a dataset into 10 equal parts are called deciles, each representing 10% of the data. There are 9 decile values from D_1 to D_9.

22. Write the formula for K^{th} decile position.

Sol. Formula: K^{th} decile position$=\dfrac{K(n+1)}{10}$

23. Find D_3 and D_7 for the following test scores: 12, 15, 20, 22, 25, 28, 30, 32, 35, 40

Sol. Here

$n = 10$

Now

$$D_3 = \frac{3(10+1)}{10} = \frac{33}{10} = 3.3$$

$\Rightarrow\ D_3 = 3^{rd}$ value$+0.3\times(4^{th}$ value-3^{rd} value)

$= 20 + 0.3 \times (22 - 20)$

$= 20 + 0.3 \times 2$

$D_3 = 20 + 0.6$

$\boxed{D_3 = 20.6}$

And

$$D_7 = \frac{7(10+1)}{10} = \frac{77}{10} = 7.7$$

$\Rightarrow\ D_7 = 7^{th}$ value$+0.7\times(8^{th}$ value -7^{th} value)

$= 30 + 0.7 \times (32 - 30)$

$D_7 = 30 + 0.7 \times 2 = 30 + 1.4$

$\Rightarrow\ \boxed{D_7 = 31.4}$

24. What are percentiles?

Ans. The values that divide an arranged dataset into 100 equal parts are called percentiles. There are 99 percentile values from P_1 to P_{99}.

25. Write formula to find position of any percentile.

Ans. Formula: position of percentile$=\dfrac{k(n+1)}{100}$

26. A student scores in 8 subjects are 55, 62, 48, 70, 85, 92, 78, 65. Find the 40^{th} percentile (P_{40}).

Sol. Arrange the data in ascending order:

48, 55, 62, 65, 70, 78, 85, 92

Here

$n = 8$

Now

$$P_{40} = \frac{40(8+1)}{100} = \frac{40(9)}{100} = \frac{360}{100}$$

$= 3.6$ (3.6^{th} position)

$\Rightarrow\ P_{40} = 3^{rd}$ value$+0.6$ (4^{th} value -3^{rd} value)

$\Rightarrow\ P_{40} = 62 + 0.6 \times (65 - 62)$

$= 62 + 0.6 \times 3 = 62 + 1.8$

$\boxed{P_{40} = 63.8}$

11.3 Scatter Diagram

27. What is a scatter diagram?

Sol. A scatter diagram (or scatter plot) is a graphical tool used to display the relationship between two numerical values (variables).

28. What is the line of best fit?

Ans. A line of best (also called trend line) is a straight line drawn through a scatter diagram that best represents the overall direction of the data.

29. What is correlation.

Sol. It is the relationship between two variables, where a change in one variable is associated with a change in the other.

30. Name three categories of correlation.

Sol. a) Positive Correlation

b) Negative Correlation

c) Zero Correlation

31. What is positive correlation?

Ans. If one variable increases as the other increases, then it is called positive correlation.

32. What is negative correlation?

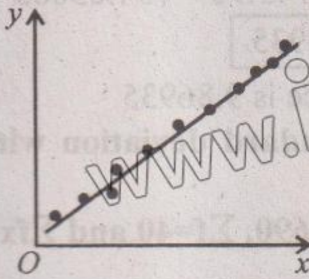
Ans. If one variable increases as the other decreases than it is called negative correlation.

33. What is zero correlation?

Sol. If there is not consistent pattern between two variables, then it is called zero correlation.

34. Plot scatter diagram of strong positive correlation.

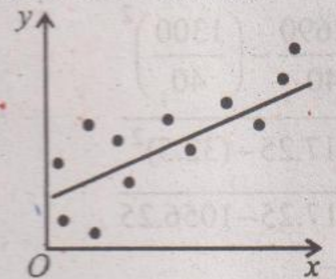
Sol.



Strong positive correlation

35. Plot scatter diagram of weak positive correlation and strong negative correlation.

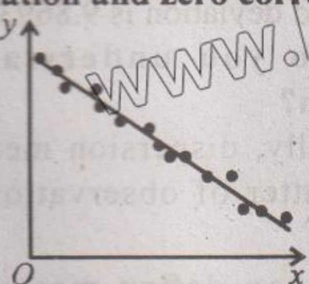
Sol.



Weak positive correlation

36. Plot scatter diagram of weak negative correlation and zero correlation.

Sol.



Strong negative correlation

37. Write the names of any four measures of dispersion.

Sol. Four measures of dispersions are:

(a) Range (b) Interquartile Range

(c) Standard Deviation (d) Variance

38. Define Range, and write its formula.

Sol. The difference between highest value (H) and lowest value (L) is called range.

Formula:

Range (R) = H – L

39. Find the range of the following data: 5, 10, 15, 20, 25, 30

Sol. Here

H = 30, L = 5

$\Rightarrow$ Range (R) = H – L

= 30 – 5

$\boxed{\text{Range (R)} = 25}$

11.4 Measures of Dispersion

40. What is variance

Sol. The average of the squares of the deviations of the values from their mean is called variance. It is denoted by σ^2 (read as sigma square).

41. Write formulas to find variance for ungrouped and grouped data.

Sol.

- **For ungrouped data**

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

or

$$\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

- **For grouped data**

$$\sigma^2 = \frac{\sum f(x-\bar{x})^2}{\sum f}$$

or

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

42. What is standard deviation?

Sol. Standard deviation is the positive square root of the mean of the squares of deviations of the given values from their mean. It is denoted by σ. (read as sigma)

43. Write the formula for S.D of ungrouped and grouped data.

Sol.

- **For ungrouped data**

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

or

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

- **For grouped data**

$$\sigma = \sqrt{\frac{\sum f(x-\overline{x})^2}{\sum f}}$$

or

$$\sigma = \sqrt{\frac{\sum fx^2}{\sum f}-\left(\frac{\sum fx}{\sum f}\right)^2}$$

44. Find variance and standard deviation if $\sum x = 63, \sum x^2 = 623\ n = 7$

Sol.

$$\sigma^2 = \frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2$$

$$\sigma^2 = \frac{623}{7}-\left(\frac{63}{7}\right)^2$$

$$= 89 - 81$$

$$\boxed{\sigma^2 = 8}$$

$$\sqrt{\sigma^2} = \sqrt{8}$$

$$\boxed{\sigma = 2.89}$$

So, variance = 8 and S.D - 2.89

45. Find variance if $\overline{x} = 9.5, \sum x^2 = 912$ and n = 8.

Sol.

$$\sigma^2 = \frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2$$

$$= \frac{912}{8}-(\overline{x})^2 \qquad \frac{\sum x}{n} = \overline{x}$$

$$= 114 - (9.5)^2 = 114 - 90.25$$

$$\boxed{\sigma^2 = 23.75}$$

So, variance is 23.75.

46. Find variance if $\sum fx = 867$, $\sum fx^2 = 23805$, and $\sum f = 32$.

Sol.

$$\sigma^2 = \frac{\sum fx^2}{\sum f}-\left(\frac{\sum fx}{\sum f}\right)^2$$

$$= \frac{23805}{32}-\left(\frac{867}{32}\right)^2$$

$$= 743.90 - (27.093)^2$$

$$= 743.90 - 734.03065$$

$$\boxed{\sigma^2 = 9.86935}$$

So, variance is 9.86935

47. Find standard deviation with given values:
$\sum fx^2 = 52690, \sum f = 40$ and $\sum fx = 1300$.

Sol.

$$\because \quad \sigma = \sqrt{\frac{\sum fx^2}{\sum f}-\left(\frac{\sum fx}{\sum f}\right)^2}$$

$$= \sqrt{\frac{52690}{40}-\left(\frac{1300}{40}\right)^2}$$

$$= \sqrt{1317.25-(32.5)^2}$$

$$= \sqrt{1317.25-1056.25}$$

$$\sigma = \sqrt{261}$$

$$\boxed{\sigma = 16.155}$$

So, standard deviation is 9.86935

48. What do you understand by dispersion?

Sol. Statistically, dispersion means the spread or scatter of observations in a dataset.

49. How do you define measures of dispersion?

Sol. The measures or techniques that are used to determine the degree or extent of variation in a data set.

50. Find range of first five multiples of 6.

Sol. First five multiples of 6 are

6, 12, 18, 24, 30

Here

H = 30, L = 6

Range (R) = H – L

= 30 – 6

$\boxed{\text{Range (R)} = 24}$