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).

INTRODUCTION

This unit focuses on essential statistical tools and techniques for analyzing and interpreting data. Students will learn to organize data using frequency tables and visualize it through cumulative frequency polygons (Ogives) and box-and-whisker plots. They will interpret key statistical measures such as median, quartiles, deciles, percentiles and interquartile range to understand data distribution. The unit also includes constructing and analyzing scatter diagrams to assess correlation and draw lines of best fit. Students will calculate and interpret measures of dispersion (range, standard deviation and variance) for grouped data and apply these to compare datasets. Real-life applications in fields such as forecasting, manufacturing, finance and economics are integrated to highlight the importance of statistical analysis in decision-making and assessing consistency in data.

In class IX, we learned that a frequency distribution organizes raw data into a table using class intervals and corresponding frequencies, allowing us to summarize and interpret large sets of data more effectively.

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	Frequencies	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	||||	4	10 + 4 = 14
65 − 69	~~||||~~ |	6	14 + 6 = 20
70 − 74	~~||||~~ ~~||||~~	10	20 + 10 = 30
75 − 79	~~||||~~ |||	8	30 + 8 = 38
80 − 84	~~||||~~	5	38 + 5 = 43
85 − 89	||||	4	43 + 4 = 47
90 − 94	||	2	47 + 2 = 49
95 − 99	|	1	49 + 1 = 50
Total		$n = \sum f = 50$	

11.1.1 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.

It helps in understanding how data values accumulate over intervals and is especially useful for estimating medians, quartiles, deciles and percentiles. The curve is typically drawn by plotting the cumulative frequencies against the upper class boundaries or lower class boundaries and joining the points with a smooth or straight-line curve. It provides a clear visual insight into the distribution and spread of the data.

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

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

11.1.2 Steps for Construction of Cumulative 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.

11.1.3 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).

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

Q_3 (Third Quartile)
75% of the data falls below this value.

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.

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.

> **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.

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 locate median, Q_1, Q_3, D_4, P_{95} and IQR on it.

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	$c.f.$
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	**$n = \Sigma f = 90$**	

Cumulative Frequency Polygon

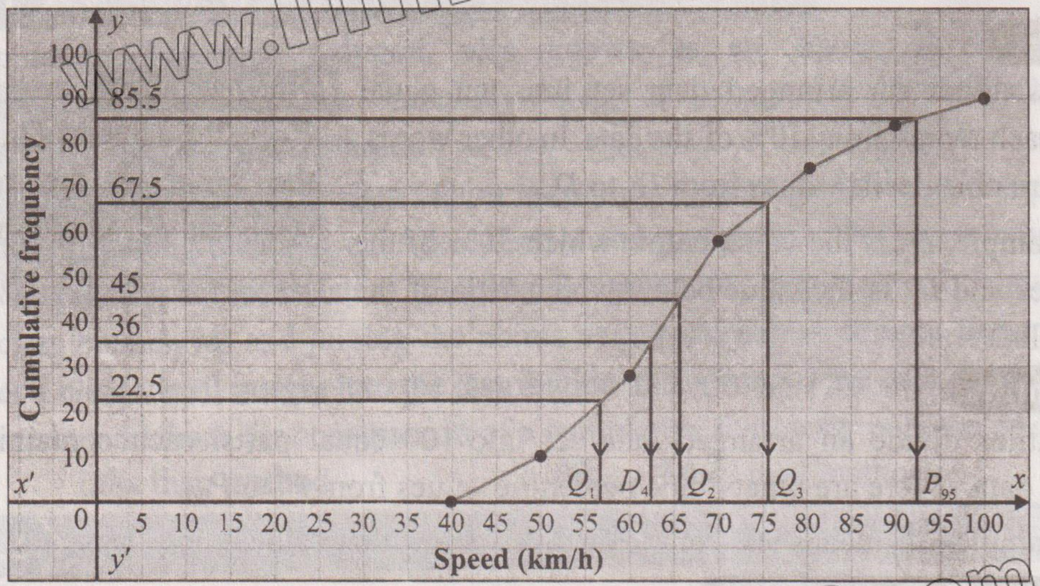


Fig. 11.1

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 $\frac{4n}{10} = \frac{4(90)}{10} = 36$. So, from the curve $D_4 = 62.7$.

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

$$IQR = Q_3 - Q_1 = 75.6 - 56.9 = 18.7$$

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. Also locate Q_2 on it.

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

Solution

Weekly wages (rupees in thousands)	No. of workers (f)	c.f.
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	$n=\sum f = 200$	

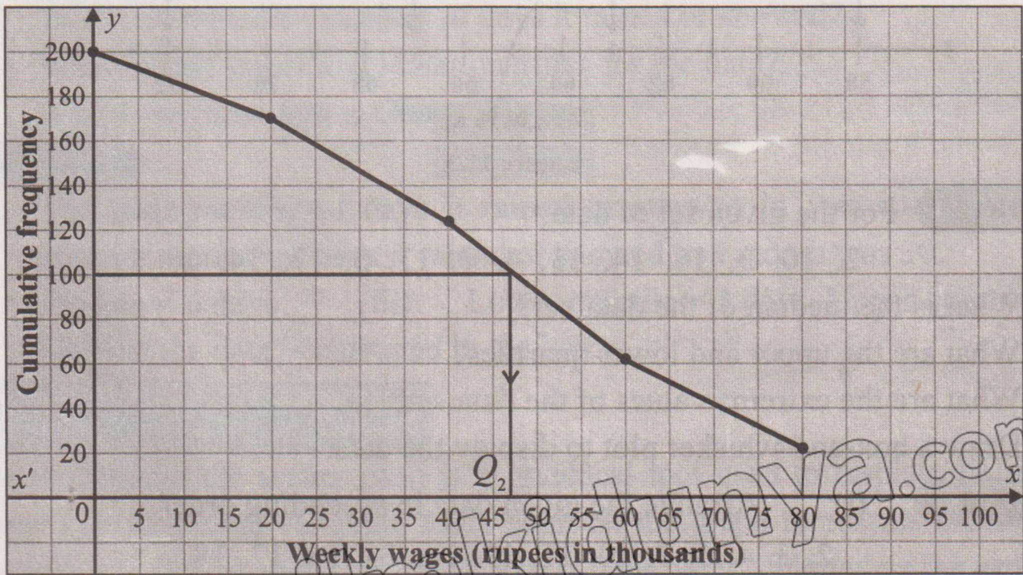


Fig. 11.2

The position of Q_2 is $\frac{n}{2} = \frac{200}{2} = 100$. So, from the curve $Q_2 = 47.2$

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 quartile (Q_1) and third quartile (Q_3).

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 value, the maximum value and the quartiles are indicated as shown in the figure given below:

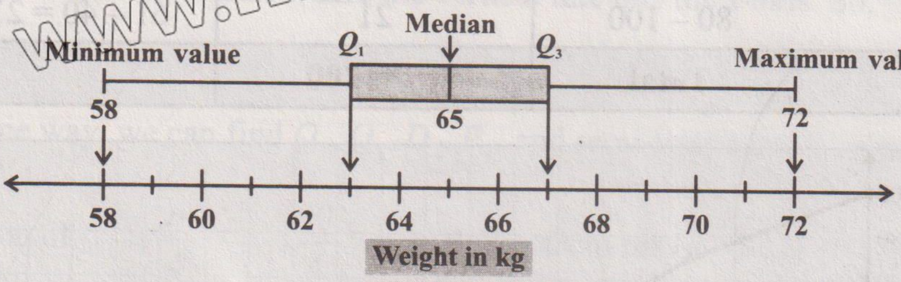


Figure 11.3

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) Lower quartile $= \left(\frac{n+1}{4}\right)^{th}$ value $= \left(\frac{13+1}{4}\right)^{th}$ value $= 3.5^{th}$

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

Lower quartile $= \frac{4+5}{2} = 4.5$

Upper quartile $= 3\left(\frac{n+1}{4}\right)^{th}$ value $= 3\left(\frac{13+1}{4}\right)^{th}$ value $= 10.5^{th}$

As the upper quartileis 10.5^{th} value, so we find average of 10^{th} and 11^{th} values.

Upper quartile $= \frac{14+16}{2} = 15$

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

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

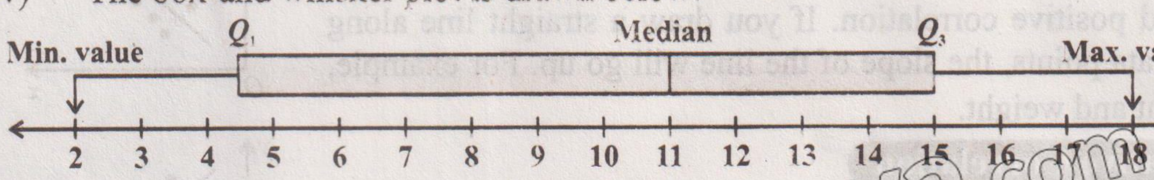
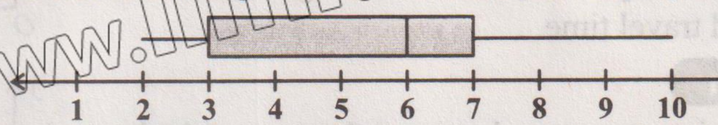


Fig. 11.4

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 tells 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.

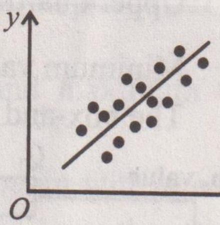
Correlation

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:

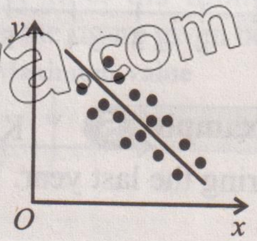
> **Do you know?**
> Scatter diagram visually shows the correlation between two variables.

Positive Correlation



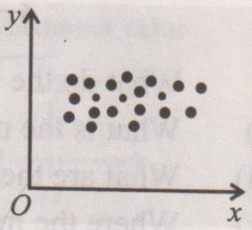
If one variable increases as the other increases, then it is called positive correlation. If you draw a straight line along the data points, the slope of the line will go up. For example, height and weight.

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.

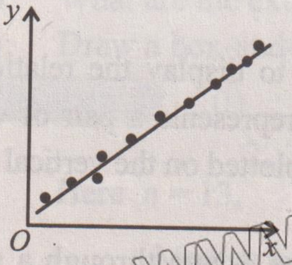
Zero Correlation



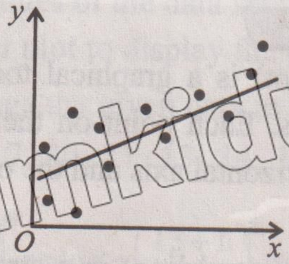
If there is not consistent pattern between the two variables, then it is called zero correlation. For example, shoe size and test scores.

Correlation Pattern of Scatter Diagram

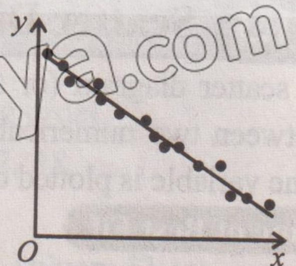
- When the value of Y increases as the X value increases, it is called strong **positive correlation**.
- When the value of Y decreases as the X value increases, it is called strong **negative correlation.**
- When the value of Y increases slightly as the X value increases, it is called **weak positive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



Strong positive correlation

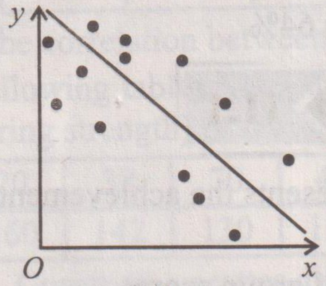


Weak positive correlation

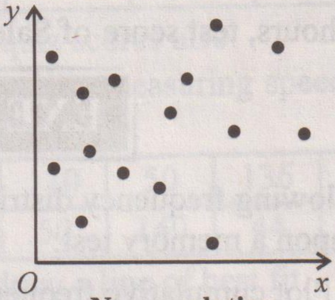


Strong negative correlation

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



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)

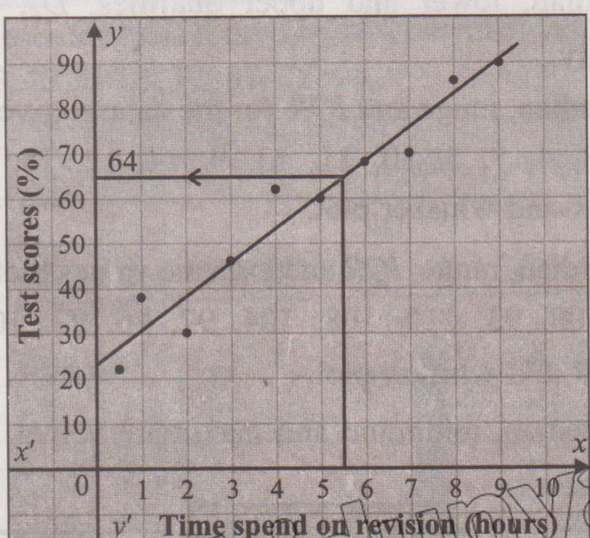


Fig. 11.5

Steps:

- Draw x-axis for "Time spend 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%.

EXERCISE 11.1

1. The following frequency 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.

   Note: Write your answers in whole numbers.

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

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

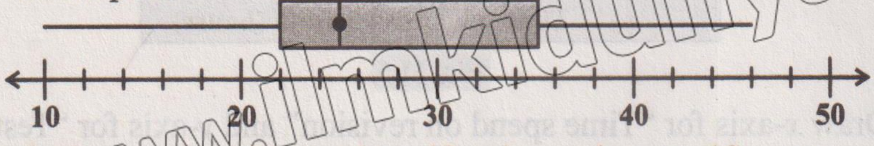
Scores	69 – 79	79 – 89	89 – 99	99 – 109	109 – 119
f	4	10	13	8	5

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

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
   Also draw a 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
   Also draw a box-and-whisker plot.

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



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

7. Construct a scatter diagram 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

Describe correlation between 1st and 2nd quiz scores also.

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 diagram and draw a line of best fit.
(ii) Describe correlation between the 1st and 2nd tests.
(iii) Abdullah scores 60 in the 2nd test. Estimate his score in the 1st test.

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.
(ii) Identify the nature of correlation between social media use and GPA.
(iii) Estimate the GPA for a student who spends 6.5 hours on social media.

11.4 Measures of Dispersion

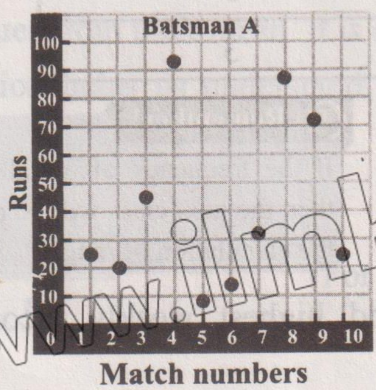
The following data shows the runs scored by two batsmen in the last 10 matches:

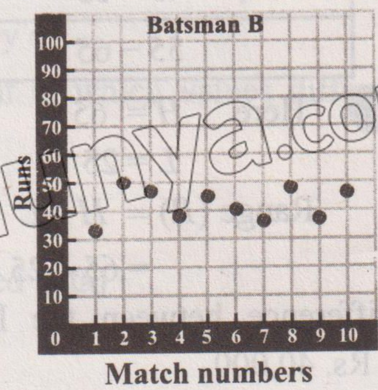
Batsman *A*: 25, 20, 45, 93, 8, 14, 32, 87, 72, 24

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 tools 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 Variations		

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

11.4.1 Range

The difference between the 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

Skilled Practice!

What is the range of first 20 prime numbers?

Solution Here $H = 30$, $L = 15$

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

$$= 30 - 15 = 15$$

Example 8 The table below shows the weekly incomes (in thousands of rupees) of a 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 $H = 65$

$L = 25$

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

$$= 65 - 25 = 40$$

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.

11.4.2 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

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

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

Let $x_1, x_2, x_3, \ldots, x_n$ be the n values with the frequencies $f_1, f_2, f_3, \ldots, f_n$ respectively.

Then,

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

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

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.

11.4.3 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.

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

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

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

$$\sigma = \sqrt{\dfrac{\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{n} f_i}}$$ (For grouped data)

Challenge!
Can variance be negative?

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

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 $\bar{x} = \dfrac{50.1 + 49.9 + 50 + 50.2 + 49.8}{5} = \dfrac{250}{5} = 50$

x	$x - \bar{x}$	$(x - \bar{x})^2$
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
		$\Sigma\left(x - \bar{x}\right)^2 = 0.10$

$$\sigma^2 = \dfrac{\Sigma\left(x - \bar{x}\right)^2}{n} = \dfrac{0.10}{5} = 0.02$$

$$\sigma = \sqrt{0.02} = 0.141\,cm$$

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	$\Sigma x^2 = 623$

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

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

Challenge! Can the standard deviation be more than the variance?

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
	$\Sigma f = 18$		$\Sigma fx = 1210$	$\Sigma fx^2 = 84450$

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

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

$$= 4691.67 - 4518.52 = 173.15$$

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

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

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

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

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

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

$$= \frac{0+16+4+4+16}{5} = \frac{40}{5} = 8$$

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

$$\bar{x}(\text{Rooma}) = \frac{\Sigma x}{n} = \frac{72+65+80+68+75}{5} = \frac{360}{5} = 72$$

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

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

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

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

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

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

(ii) $\sigma(\text{Ali}) = 2.83$, $\sigma(\text{Rooma}) = 5.25$

Hence, Ali was more consistent.

EXERCISE 11.2

1. Find the range of the following data sets:

   (i) 63, 89, 98, 125, 79, 108, 117, 60 (ii) 43.5, 13.6, 18.9, 38.4, 61.4, 29.4

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.

3. Calculate the range of the following data:

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

4. A group of 7 workers reported the number of items they assembled in a day as:

52, 55, 50, 53, 54, 56, 52

Find the standard deviation and variance of the items assembled.

5. A librarian recorded the number of visitors during 5 days of a week.

120, 135, 130, 125, 140

Calculate the variance and standard deviation of visitors.

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

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

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 consistent?

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 consistency.

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.

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

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

13. Find variance and standard deviation:

(i) $\Sigma x = 45, \Sigma x^2 = 421, n = 5$ (ii) $\Sigma fx = 210, \Sigma fx^2 = 7560, \Sigma f = 6$

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

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

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 locate median, Q_1, Q_3, D_7, P_{90} and interquartile range on it.

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

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

Exercise 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.

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

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

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.

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?

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.