

13

STATISTICAL DIAGRAMS

LEARNING OBJECTIVES

In this topic, we will learn to:

- interpret and analyse dot diagrams
- interpret and analyse histograms
- interpret and analyse stem-and-leaf diagrams

13.1 DOT DIAGRAMS

1. A dot diagram consists of a horizontal number line in which each value on the number line is denoted by a dot above it.
2. The number of dots above each value on the number line shows the frequency of the value.
3. Disadvantages:
 - (a) It is not suitable when there is a large data set.
 - (b) It is not suitable when a data set has a large range.

Note:

A dot diagram is similar to a bar chart, just that it is expressed in a different format.

WORKED EXAMPLE 1

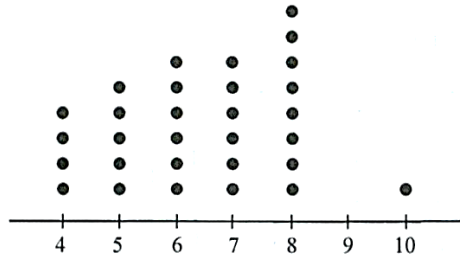
30 students in a school were asked how many hours they slept last night. The results are given below.

7	6	5	5	4	8
5	4	6	7	5	7
8	8	8	8	8	7
4	7	6	7	8	6
5	10	6	8	6	4

- (a) Represent the data using a dot diagram.
- (b) Find the number of hours most of the students slept last night,
- (c) Find the percentage of students who slept less than 5 hours last night.
- (d) Describe briefly the distribution of the data.

Worked Solution:

(a)



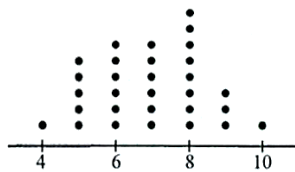
(b) Number of hours most of the students slept last night = **8 hours**

(c) Percentage of students who slept less than 5 hours last night = $\frac{4}{30} \times 100\%$
 $= 13\frac{1}{3}\%$

(d) The numbers of hours slept for the students range from 4 hours to 10 hours. The numbers of hours slept cluster around 4 hours to 8 hours, with an extreme value of 10 hours. The distribution is not symmetrical.

Student's common mistake:

In the example, it is wrong to construct the dot diagram as shown below.



Dots must be placed above each corresponding value on the number line.

13.2 HISTOGRAMS FOR UNGROUPED DATA

- When we draw a rectangle over each column of dots in a dot diagram, a histogram for ungrouped data is obtained.
- Disadvantages:
 - It is more abstract compared to a dot diagram as frequency is represented by the area of a column.
 - A data may be misinterpreted if the frequency does not start from zero.

WORKED EXAMPLE 2

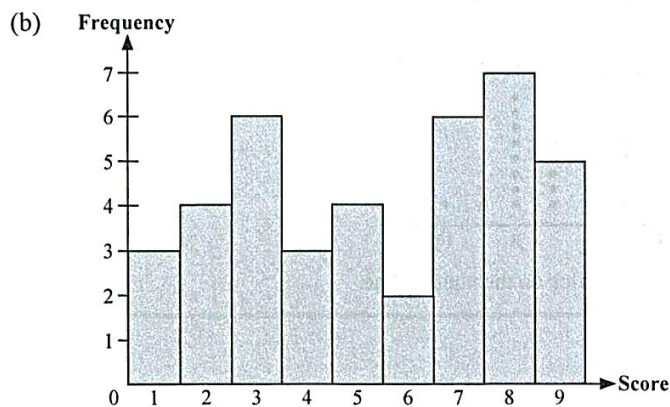
A class of 40 children took a quiz. Their scores are given below.

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

- Construct a frequency table for the data.
- Draw a histogram to illustrate the data.
- Find the most common score.
- Find the fraction of students with a score of more than 5.

Worked Solution:

- | Score | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-----------|---|---|---|---|---|---|---|---|---|
| Frequency | 3 | 4 | 6 | 3 | 4 | 2 | 6 | 7 | 5 |



Note:

In a histogram, there is no space between two adjacent columns.

- Most common score = 8
- Number of students with a score of more than 5

$$= 2 + 6 + 7 + 5$$

$$= 20$$

Fraction of students with a score of more than 5

$$= \frac{20}{40}$$

$$= \frac{1}{2}$$

13.3 STEM-AND-LEAF DIAGRAMS

1. A stem-and-leaf diagram displays a set of data with many different values that can be grouped into class intervals and is used when we want to retain individual data values.
2. In a stem-and-leaf diagram, each data is organised in two sections, namely the stem and the leaf. The more significant digit(s) of each data value is organised in the 'stem' column. The next significant digit of each data value is displayed in the 'leaf' column.
3. Disadvantages:
 - (a) It is not suitable when there is a large data set.
 - (b) It is not suitable when a data set has too small or too large a range.

WORKED EXAMPLE 3

The masses, in kilograms, of 25 students were measured. The results are given below.

43	50	48	56	65
70	45	52	55	56
55	49	57	62	64
54	48	64	68	75
52	57	59	45	52

- Represent the data using a stem-and-leaf diagram.
- Find the mass of the heaviest student.
- Find the class interval with the highest frequency.
- Find the percentage of students whose masses are more than 60 kg.

Worked Solution:

- (a)

[illegible]

Key: 4 | 3 means 43 kg

Note:

The data are grouped into equal class intervals by the stems in the stem-and-leaf diagram. The stems are arranged in ascending order in a vertical column. The leaf of each value is in ascending order in the row corresponding to its stem. The key for interpreting the diagram is given.

- (b) Mass of heaviest student = **75 kg**
- (c) Class interval with the highest frequency = **50 to 59**
- (d) Percentage of students whose masses are more than 60 kg
 $= \frac{7}{25} \times 100\%$
 $= \mathbf{28\%}$

4. In a stem-and-leaf diagram with split stems, the range of the stems (particularly for diagrams with too few stems but fairly long leaves) are broken into two halves so that it will be easier to interpret the diagram.
5. The first half is for the leaves 0 to 4 and the second half is for the leaves 5 to 9.

WORKED EXAMPLE 4

- (a) Represent the stem-and-leaf diagram in Example 3 using a stem-and-leaf diagram with split stems.
- (b) Describe briefly the shape of this distribution.
- (c) Which stem-and-leaf diagram is more suitable to represent the data? Explain your answer.

Worked Solution:

(a)

4	3
4	5 5 8 8 9
5	0 2 2 2 4
5	5 5 6 6 7 7 9
6	2 4 4
6	5 8
7	0
7	5

Key: 4 | 3 means 43 kg

- (b) The masses of the students cluster around 45 kg to 59 kg and the distribution is with a thinner tail towards the higher mass.
- (c) The stem-and-leaf diagram in Example 3 is more suitable to represent this data since it shows a more symmetrical distribution.

6. A back-to-back stem-and-leaf diagram is used to compare two sets of related data.
7. It has a common stem in the middle and the leaves of each data set on its two sides.

WORKED EXAMPLE 5

The heights, in centimetres, of 10-year-old children in Town A and Town B are measured. The results are given below.

Town A	141	138	146	140	131	138	148	136	128	135
	133	142	145	125	128	144	147	132	135	130
Town B	135	120	125	122	130	126	140	133	145	128
	125	123	134	131	124	142	127	136	122	121

- (a) Represent the two sets of data using a back-to-back stem-and-leaf diagram.
- (b) In which town are the children generally taller? Explain your answer.

Worked Solution:

(a)

Town A					Town B						
8					12	0	1	2	2	3	4
		8	8	5	12	5	5	6	7	8	
	3	2	1	0	13	0	1	3	4		
	8	6	5	5	13	5	6				
	4	2	1	0	14	0	2				
	8	7	6	5	14	5					
Key (Town A)					Key (Town B)						
5 12 means 125 cm					12 0 means 120 cm						

- (b) The distribution of the heights of the children in Town A is fairly symmetrical with a peak at the range of 135 cm to 138 cm. However, the heights of the children in Town B cluster around 120 cm to 134 cm and the distribution is with a thinner tail towards the greater values of heights. In general, the children in Town A are taller than the children in Town B.

13.4 HISTOGRAMS FOR GROUPED DATA

1. A histogram for grouped data displays a set of data with many different values that can be grouped into class intervals and is used when we do not need to retain individual data values.
2. Disadvantages:
 - (a) The original data values are lost in a histogram.
 - (b) It is not possible to compare two sets of data in a histogram.

WORKED EXAMPLE 6

The average speed, x km/h, of each of 30 dogs in a dog racing competition was recorded. The results are shown below.

77	75	73	74	83	84	82	81	86	80
77	69	66	85	81	82	84	81	70	72
70	69	62	65	76	73	80	63	72	83

- (a) Construct a frequency table using class intervals $60 < x \leq 65$, $65 < x \leq 70$ and so on for the data.
- (b) Draw a histogram to illustrate the data.
- (c) Find the percentage of dogs that run at an average speed of not more than 80 km/h.

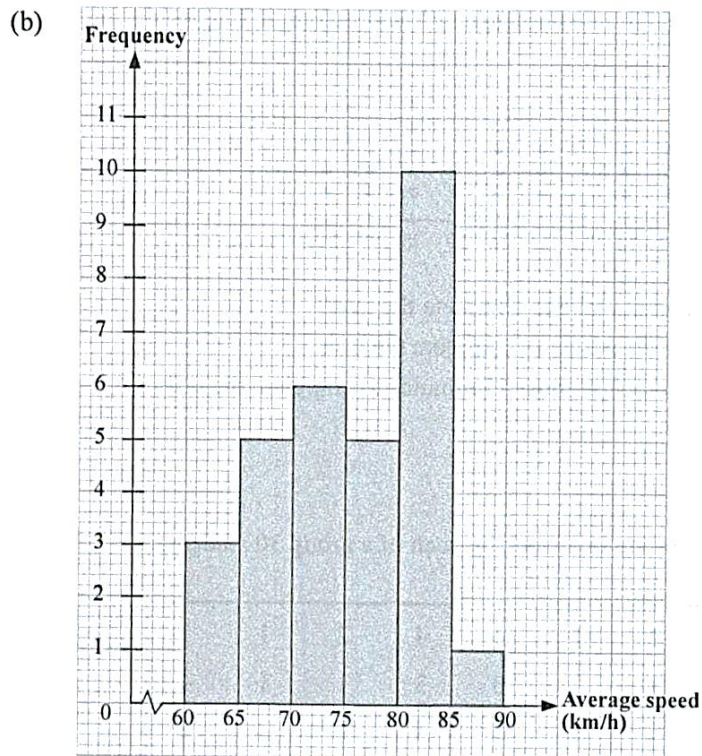
Worked Solution:

(a)

Average speed (x km/h)	Tally	Frequency
$60 < x \leq 65$		3
$65 < x \leq 70$	++++	5
$70 < x \leq 75$	++++	6
$75 < x \leq 80$	++++	5
$80 < x \leq 85$	++++ ++	10
$85 < x \leq 90$		1
Total		30

Note:

The class interval $60 < x \leq 65$ includes 65 but excludes 60.



- (c) Number of dogs that run at an average speed of not more than 80 km/h
 $= 3 + 5 + 6 + 5$
 $= 19$
 Percentage of dogs that run at an average speed of not more than 80 km/h
 $= \frac{19}{30} \times 100\%$
 $= 63\frac{1}{3}\%$

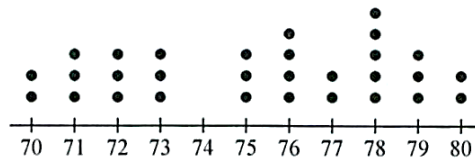
PRACTICE QUESTIONS

1. The table shows the number of letters received in a month by each of a group 20 adults.

17	10	13	15	16	14	14	12	15	15
17	19	14	15	12	14	17	16	16	14

- Represent the data on a dot diagram.
- Find the most common number of letters received in the month.
- Find the percentage of adults who received at least 17 letters in the month.
- Describe briefly the distribution of the data.

2. The time taken, in minutes, for each adult to complete a task was recorded. The results are shown in the dot diagram.



- (a) How many minutes did most adults take to complete the task?
 (b) Find the longest time taken for the adults to complete the task.
 (c) Find the fraction of adults who took at most 75 minutes to complete the task.
3. The table shows the number of apples ate in a week by each of a group 30 students.

3	5	3	2	1	4	4	2	1	1
4	3	2	3	2	0	3	1	1	4
1	3	2	5	5	3	4	2	3	2

- (a) Construct a frequency table for the data.
 (b) Draw a histogram to illustrate the data.
 (c) Find the most common number of apples ate in a week.
4. The table shows the heights, in centimetres, of 30 beauty contestants.

170	184	177	169	174	181	175	180	186	168
178	174	159	179	182	169	181	175	170	181
168	181	190	183	175	191	159	171	192	176

- (a) Represent the data using a stem-and-leaf diagram.
 (b) Find the height of the tallest beauty contestant.
 (c) Find the most common height.
 (d) Find the percentage of beauty contestants that are shorter than 172 cm.

5. The average amount of money, in dollars, saved by each of 40 adults in a month was recorded. The results are shown in the stem-and-leaf diagram.

31	4					
31	6	9				
32	0	1	4	4		
32	5	5	6	8	8	9
33	1	1	2			
33	6	6	7	8		
34	0	3				
34	5	6	7	8		
35	0	1	2	2		
35	6	7	8	9		
36	0	4	4			
36	7	8	8			

Key: 31 | 4 means \$314

- (a) State the class interval with the most data value.
 (b) Find the percentage of adults that saved less than \$340 in a month.
 (c) An adult is selected at random. Given that the probability that the adult saved at least \$ x a month is $\frac{3}{20}$, find the value of x .
6. The masses, in kilograms, of two groups of students are recorded in the following table.

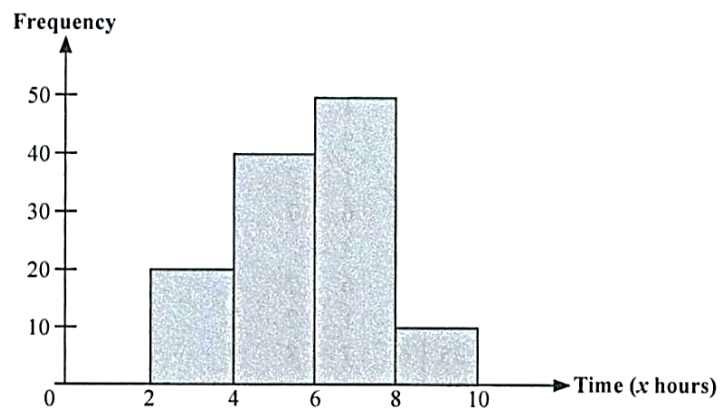
Group A	43	45	49	40	53	52
	50	40	55	61	59	44
Group B	59	67	53	65	62	54
	50	48	59	44	60	57

- (a) Represent the two sets of data using a back-to-back stem-and-leaf diagram.
 (b) Which class interval for each group has the fewest students?
 (c) Which group of students is heavier? Explain your answer.
7. The table shows the marks, x , obtained by 30 students in a Mathematics test.

51	46	59	57	47	50	61	55	56	54
48	44	74	63	46	45	58	52	48	66
67	53	57	53	44	41	70	51	56	43

- (a) Construct a frequency table using class intervals $40 < x \leq 45$, $45 < x \leq 50$ and so on for the data.
 (b) Draw a histogram to illustrate the data.
 (c) Find the percentage of students who obtained more than 60 marks.

8. The histogram shows the times, x hours, taken by the students in a school to complete a project.



- (a) Complete the following grouped frequency table for the data.

Time (x hours)	Frequency
$0 \leq x < 2$	
$2 \leq x < 4$	
$4 \leq x < 6$	
$6 \leq x < 8$	
$8 \leq x < 10$	

- (b) State the class interval containing the most students.
(c) A student is picked at random. Find the probability that the student took at least 4 hours but less than 8 hours to complete the project.