

5 Data Analysis

5.1 Frequency Tables: Discrete Ungrouped Data

In this section we revise collecting data to draw vertical line graphs or pie charts.



Example 1

The pupils in Mr Middleton's class take a maths test and get scores out of 10, which are listed below:

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

Illustrate these results using a pie chart.



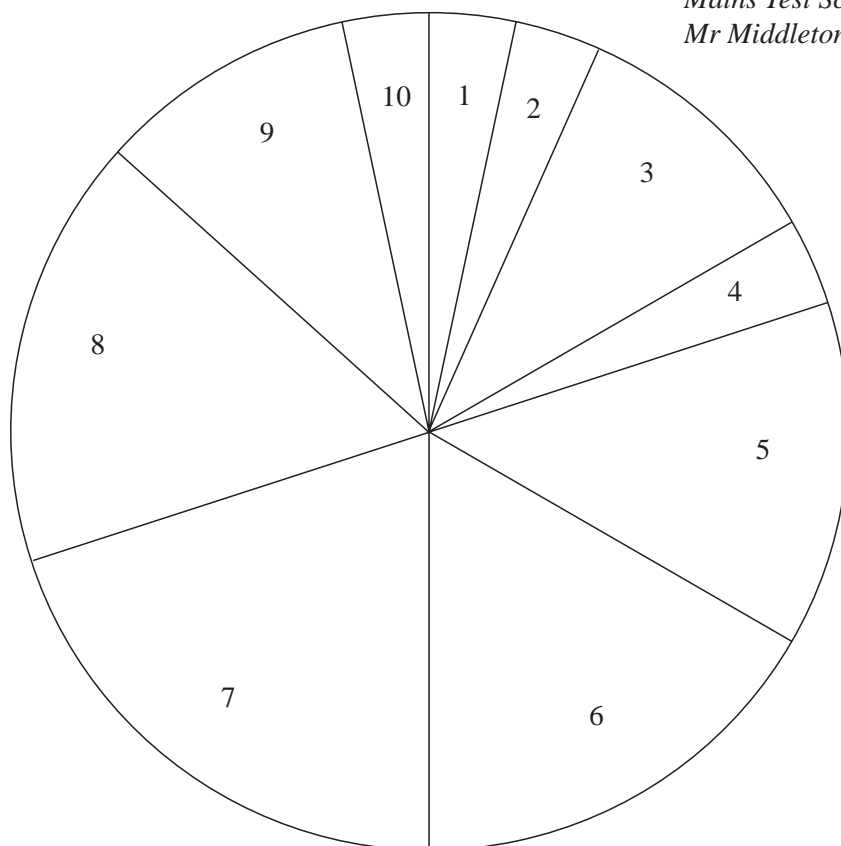
Solution

First construct a tally chart and then calculate the angles for the pie chart.

Score	Tally	Frequency	Angle
1		1	$\frac{360}{30} \times 1 = 12^\circ$
2		1	$\frac{360}{30} \times 1 = 12^\circ$
3		3	$\frac{360}{30} \times 3 = 36^\circ$
4		1	$\frac{360}{30} \times 1 = 12^\circ$
5		4	$\frac{360}{30} \times 4 = 48^\circ$
6		5	$\frac{360}{30} \times 5 = 60^\circ$
7		6	$\frac{360}{30} \times 6 = 72^\circ$
8		5	$\frac{360}{30} \times 5 = 60^\circ$
9		3	$\frac{360}{30} \times 3 = 36^\circ$
10		1	$\frac{360}{30} \times 1 = 12^\circ$
		TOTAL 30	360°

The pie chart can now be drawn as shown below:

*Pie Chart to Show
Maths Test Scores for
Mr Middleton's Class*



Note: Remember to always give a title for the pie chart.



Example 2

Mrs Panni gave her class the same maths test as Mr Middleton's class in Example 1. The scores for her class are given below:

3	4	4	4	5	6	7	7	4	3
5	7	6	7	5	8	3	5	6	4
6	7	7	6	6	4	6	5	6	7

Draw vertical line graphs for *both* classes and comment on the differences that they show.

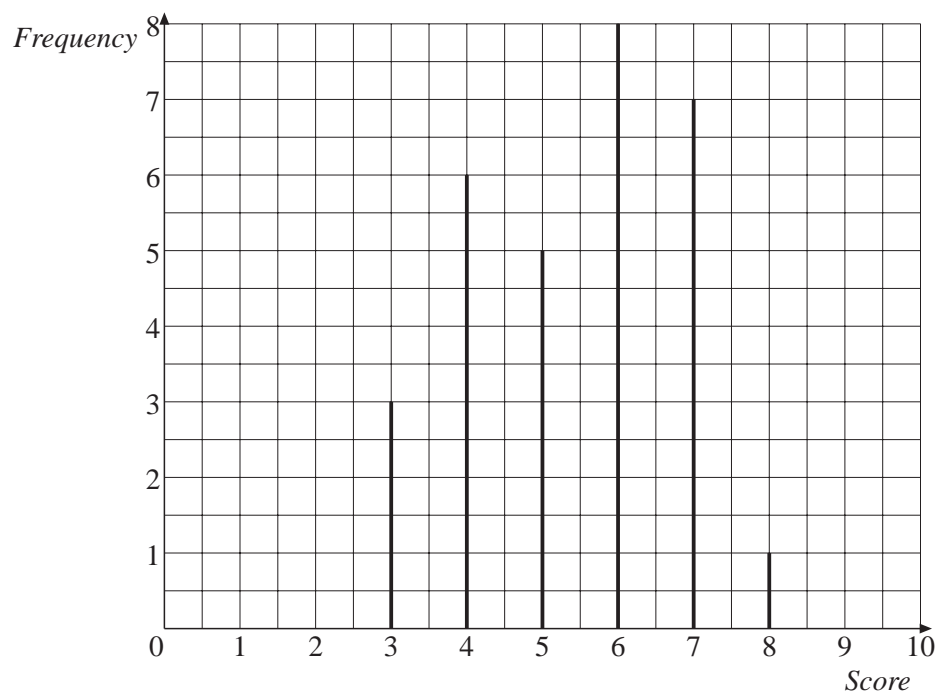


Solution

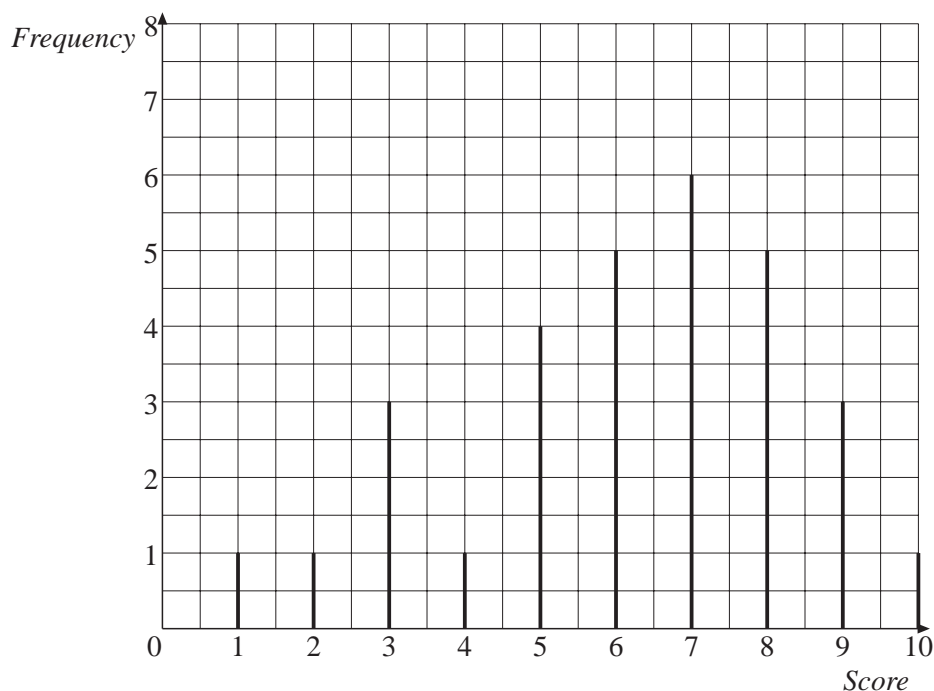
First construct a tally chart for Mrs Panni's class.

<i>Score</i>	<i>Tally</i>	<i>Frequency</i>
1		0
2		0
3		3
4		6
5		5
6		8
7		7
8		1
9		0
10		0

The vertical line graph for Mrs Panni's class is shown below:



The vertical line graph for Mr Middleton's class is shown below:



From the graphs we can see that Mr Middleton's class has some pupils who did better than those in Mrs Panni's class, but it also has some pupils who had lower scores.



Exercises

- Emma stops at the park on her way to school every day for 2 weeks to look for conkers. The number of conkers she finds each day is listed below:

2 10 1 2 3 4 5 3 2 4

 - Draw a pie chart to illustrate these data.
 - Can you think of a possible reason for the unusual number in the data?
- Mr Rafiq runs a video library. Over a period of 3 days he notes how many videos have been borrowed during each hour. His records are shown below:

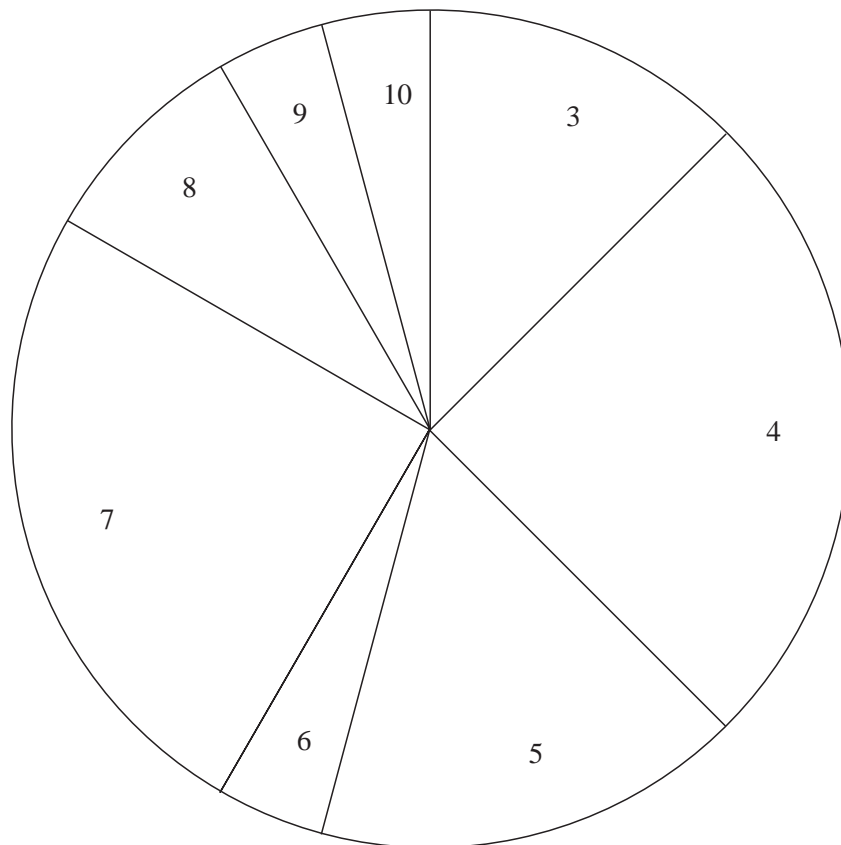
3	2	4	3	8	7	9	3	10	2
4	6	8	4	12	10	8	5	6	4
3	2	6	9	4	5	7	6	5	7

 - Draw a pie chart to illustrate these data.
 - What was the *largest* number of videos hired in any hour?
 - What was the *least* number of videos hired in any hour?
 - What was the *most common* number of videos hired in any hour?

3. The total number of goals scored in each of the Premier League matches one Saturday were:

0 1 0 4 5 4 2 3 1 4

- (a) Illustrate these data on a pie chart.
 (b) Which number of goals was the most common?
4. The pie chart below illustrates the scores obtained on a maths test by a class of 24 pupils:



- (a) How many degrees represent each pupil?
 (b) How many pupils get each score?
5. The National Curriculum levels reached in maths by a class at Key Stage 3 were:

6 5 6 4 3 4 4 5 6 4 5 6 6 5
 5 6 4 4 6 6 5 5 6 4 6 5 4 5

- (a) Draw a vertical line diagram for this data.
 (b) Which level was obtained by most pupils?

6. The class in question 5 were also given National Curriculum levels for English and these are listed below:

6	4	5	3	4	3	4	6	5	4	6	5	5	6
4	6	4	5	5	5	4	5	5	3	6	4	4	4

- (a) Draw a vertical line graph for the data.
- (b) Compare this graph with the one for question 5 and comment on any differences.
7. A calculator was used to produce 40 random digits. These are listed below:

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

- (a) Draw a vertical line graph to illustrate this data.
- (b) Do you think that the calculator is good at producing random numbers? How does the graph support your answer?
8. A train company keeps a daily record of the number of trains that were late for two different months and these data are shown below:

<i>January</i>	1	0	2	7	3	1	2	4	5	1	2
	4	7	4	3	0	1	4	1	3	2	1
	2	3	4	5	6	0	1	2	9		
<i>February</i>	5	6	5	4	3	7	8	1	4	3	4
	5	6	4	3	4	2	6	4	8	9	5
	3	4	3	5	4	4					

- (a) Draw a vertical line graph for each month.
- (b) In which month was the better service provided to passengers?
- (c) Why is the amount of data different for each of the two months?
- (d) If January had the same number of days as February, would your answer to (b) be the same?

9. Peter grows tomatoes in his greenhouse. During August he keeps records of the number of tomatoes he picks each day. The results for 1997 and 1998 are shown below:

1997	6	7	8	10	12	14	10	15	12	9	6
	8	10	8	7	9	16	20	19	12	10	11
	12	10	9	8	7	9	11	12	8		
1998	11	11	14	16	15	12	12	11	8	9	20
	19	17	18	19	15	16	17	15	12	11	10
	8	17	18	20	15	16	17	19	15		

- (a) Draw vertical line graphs for each year.
 (b) Describe how the two years compare.

5.2 Mean, Median, Mode and Range

The *mean*, *median* and *mode* are types of average.

The *range* gives a measure of the *spread* of a set of data.

Definition

Example

$$\text{Mean} = \frac{\text{sum of all data}}{\text{number of values}}$$

For 1, 2, 2, 3, 4

$$\begin{aligned}\text{Mean} &= \frac{1 + 2 + 2 + 3 + 4}{5} \\ &= \frac{12}{5} \\ &= 2.4\end{aligned}$$

Mode = most common value

For 1, 2, 2, 3, 4

$$\text{Mode} = 2$$

For 1, 2, 2, 3, 4, 4, 5

$$\text{Mode} = 2 \text{ and } 4$$

Median = middle value when data is arranged in order

For 1, 2, (2), 3, 4

$$\text{Median} = 2$$

For 1, 2, (2, 3), 4, 4

$$\begin{aligned}\text{Median} &= \frac{2 + 3}{2} \\ &= 2.5\end{aligned}$$

Definition

Range = largest value – smallest value

Example

For 1, 2, 2, 3, 4

$$\begin{aligned} \text{Range} &= 4 - 1 \\ &= 3 \end{aligned}$$

The *mean*, *median*, *mode* and *range* can also be calculated when the data is presented in the form of a frequency table.

**Example 1**

For the data presented in the table opposite, calculate:

- (a) the *mode*,
- (b) the *range*,
- (c) the *median*,
- (d) the *mean*.

<i>Score</i>	<i>Frequency</i>
0	2
1	6
2	12
3	4
4	1

**Solution**

- (a) The mode is the most common score. In this case,
mode = 2
- (b) Largest score = 4, smallest score = 0,
range = 4 – 0
= 4
- (c) The median is the middle value. As there are 25 scores, the middle value is the 13th score (12 above and 12 below).

When in order:

the first 2 values are 0,
the next 6 values are 1,
therefore the 3rd to 8th values are 1.

The next 12 values are 2,
therefore the 9th to 20th values are 2.

The 13th value is in this group,
so the 13th value is 2.

So the median = 2

- (d) To calculate the mean, complete a table like the one below:

<i>Score</i>	<i>Frequency</i>	<i>Score × Frequency</i>	
0	2	$0 \times 2 = 0$	
1	6	$1 \times 6 = 6$	
2	12	$2 \times 12 = 24$	Twelve pupils scored 2, a total of 24
3	4	$3 \times 4 = 12$	
4	1	$4 \times 1 = 4$	
Totals	25	46	

$$\text{Mean} = \frac{46}{25}$$

$$= 1.84$$

Total number of pupils

Total of the scores



Example 2

Calculate the *mean* and *median* for the data in the table opposite.

<i>Price</i>	<i>Frequency</i>
30p	1
31p	3
32p	4
33p	8
34p	2
35p	2



Solution

- (a) To calculate the mean, construct a table like the one below:

<i>Price (pence)</i>	<i>Frequency</i>	<i>Price × Frequency</i>
30	1	$30 \times 1 = 30$
31	3	$31 \times 3 = 93$
32	4	$32 \times 4 = 128$
33	8	$33 \times 8 = 264$
34	2	$34 \times 2 = 68$
35	2	$35 \times 2 = 70$
Totals	20	653

$$\text{Mean} = \frac{653}{20}$$

$$= 32.65\text{p}$$

As there are 20 values, the median will be between the 10th and 11th values.

When in order:

1st value is 30p

Next 3 values are 31p

So the 2nd to 4th values are 31p.

Next 4 values are 32p

So the 5th to 8th values are 32p

Next 8 values are 33p

So 9th to 17th values are 33p.

So both the 10th and 11th values are 33p.

Median = 33p

Note: If the 10th and 11th values had been different from one another, we would have used a value halfway between them (the mean of the two numbers).

The symbol Σ (Greek 'sigma') means 'the sum of' or 'the total of'.

So,

$$\text{mean} = \frac{\Sigma \text{ frequency} \times \text{value}}{\Sigma \text{ frequency}}$$

Sometimes we use f to stand for frequency, x for the values and \bar{x} for the mean, so that,

$$\bar{x} = \frac{\Sigma f x}{\Sigma f}$$



Exercises

1. Calculate the *mean*, *median*, *mode* and *range* for each set of data below:

(a) 3, 6, 3, 7, 4, 3, 9

(b) 11, 10, 12, 12, 9, 10, 14, 12, 9

(c) 6, 9, 10, 7, 8, 5

(d) 2, 9, 7, 3, 5, 5, 6, 5, 4, 9.

2. (a) Copy and complete the table below:

<i>Score</i>	<i>Frequency</i>	<i>Score \times Frequency</i>
0	2	
1	6	
2	8	
3	3	
4	0	
5	1	
Totals		

- (b) Calculate the mean score.

3. The number of goals scored by a hockey team in each match of a season is listed below:

6	2	2	3	1	0	0	1	2	3
3	5	4	2	0	1	1	0	2	1
1	1	1	2	1	1	0	2	3	4

- (a) Copy and complete the table below:

<i>No. of Goals</i>	<i>Tally</i>	<i>Frequency</i>	<i>No. of Goals \times Frequency</i>
0			
1			
2			
3			
4			
5			
6			

- (b) Calculate the *mean*.
 (c) Calculate the *median*.
 (d) What is the *mode*?
 (e) What is the *range*?

4. The price of a litre of petrol at some garages was recorded and the results are given in the table opposite:

Calculate the *mean*, *median* and *mode* of these data.

<i>Price</i>	<i>Frequency</i>
74p	1
75p	2
76p	8
77p	10
78p	2
79p	1
80p	1

5. A class collected data on the number of children in their families, and this information is listed below:

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

- (a) Calculate the *mean* number of children per family.
 (b) Calculate the *median* number of children per family.
 (c) Why are there no 'zeros' ?

6. Professor Baker keeps a record of his golf scores, as shown in the table opposite:

Calculate his mean score.

<i>Golf Score</i>	<i>Frequency</i>
70	3
71	4
72	4
73	4
74	3
75	2

7. A class collected data on their shoe sizes and presented it in the table opposite:

- (a) Calculate the *mean*, *median* and *mode* for the data.
 (b) Which of the three types of average is the most useful to a shoe shop manager ordering stock?

<i>Shoe Size</i>	<i>Frequency</i>
3	2
4	7
5	6
6	5
7	3
8	2

8. Fataimenta sells vacuum cleaners and the table shows how many she sells each day in a 25-day period.

<i>No. Sold per Day</i>	<i>Frequency</i>
0	2
1	6
2	6
3	2
4	8
5	1

- (a) Calculate the *mean*, *median* and *mode* for the data.
- (b) Which of the averages gives the *best* impression of her sales figures?

9. Classes 8A and 8B have a sponsored spelling competition. The tables below give the number of correct spellings for both classes.

Class A

<i>No. of Correct Spellings</i>	<i>Frequency</i>
1	0
2	1
3	1
4	1
5	1
6	6
7	5
8	6
9	1
10	3

Class B

<i>No. of Correct Spellings</i>	<i>Frequency</i>
1	0
2	1
3	1
4	5
5	2
6	3
7	1
8	2
9	4
10	6

- (a) Calculate the *mean*, *median* and *mode* for each class.
- (b) Which average makes class A appear to be better at spelling?
- (c) Which average makes class B appear to be better at spelling?
10. Paul and David play golf. The scores for their last 20 matches are given below:

<i>Score</i>	67	68	69	70	71	72	73	74	75	76	77	78
<i>Paul's Frequency</i>	1	1	1	1	2	5	1	2	3	1	1	1
<i>David's Frequency</i>	0	0	3	6	2	0	0	0	0	2	4	3

Produce arguments, supported by an average for each player, to show that each could be considered the better player.