■ Mixed Practice

1 A librarian is investigating the number of books borrowed from the school library over a period of 10 weeks. She decided to select a sample of 10 days and record the number of books borrowed on that day.

a She first suggests selecting a day at random and then selecting every seventh day after that.
   i State the name of this sampling technique.
   ii Identify one possible source of bias in this sample.

b The librarian changes her mind and selects a simple random sample of 10 days instead.
   i Explain what is meant by a simple random sample in this context.
   ii State one advantage of a simple random sample compared to the sampling method from part **a**.

c For the days in the sample, the numbers of books borrowed were:
17, 16, 21, 16, 19, 20, 18, 11, 22, 14
Find
   i the range of the data
   ii the mean number of books borrowed per day
   iii the standard deviation of the data.

2 The table shows the maximum temperature ($T°C$) and the number of cold drinks (n) sold by a small shop on a random sample of nine summer days.

T	21	28	19	21	32	22	27	18	30
n	20	37	21	18	35	25	31	17	38

a Using technology, or otherwise, plot the data on a scatter graph.
b Describe the relationship between the temperature and the sales of cold drinks.
c Find the equation of the regression line of n on T.
d Use your regression line to estimate the number of cold drinks sold on the day when the maximum temperature is $26°C$.

3 The masses of 50 cats are summarized in the grouped frequency table:

Mass (kg)	$1.2 \leqslant m < 1.6$	$1.6 \leqslant m < 2.0$	$2.0 \leqslant m < 2.4$	$2.4 \leqslant m < 2.8$	$2.8 \leqslant m < 3.2$
Frequency	4	10	8	16	12

a Use this table to estimate the mean mass of a cat in this sample. Explain why your answer is only an estimate.
b Use technology to create a cumulative frequency graph.
c Use your graph to find the median and the interquartile range of the masses.
d Create a box plot to represent the data. You may assume that there are no outliers.

4 A survey was carried out on a road to determine the number of passengers in each car (excluding the driver). The table shows the results of the survey.

Number of passengers	0	1	2	3	4
Number of cars	37	23	36	15	9

a State whether the data are discrete or continuous.
b Write down the mode.
c Use your GDC to find
   i the mean number of passengers per car
   ii the median number of passengers per car
   **iii** the standard deviation.

5 Two groups of 40 students were asked how many books they have read in the last two months. The results for **the first group** are shown in the following table.

Number of books read	Frequency
2	5
3	8
4	13
5	7
6	4
7	2
8	1

The quartiles for these results are 3 and 5.
a Write down the value of the median for these results.
b Draw a box-and-whisker diagram for these results.

The results for **the second group** of 40 students are shown in the following box-and-whisker diagram.

Number of books read
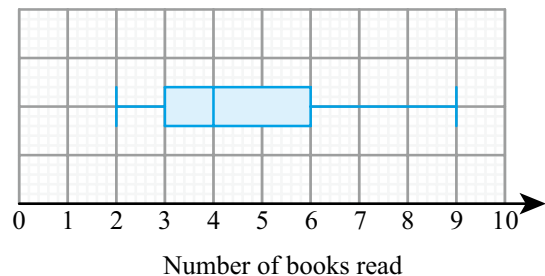
c Estimate the number of students **in the second group** who have read at least 6 books.

6 The following table shows the Diploma score x and university entrance mark y for seven IB Diploma students.

Diploma score (x)	28	30	27	31	32	25	27
University entrance mark (y)	73.9	78.1	70.2	82.2	85.5	62.7	69.4

a Find the correlation coefficient.

The relationship can be modelled by the regression line with equation $y = ax + b$.
b Write down the value of a and b.

Rita scored a total of 26 in her IB Diploma.
c Use your regression line to estimate Rita's university entrance mark.

7 A student recorded, over a period of several months, the amount of time he waited in the queue for lunch at the college canteen. He summarized the results in this cumulative frequency table.

Time (minutes)	$\leqslant 2$	$\leqslant 4$	$\leqslant 6$	$\leqslant 8$	$\leqslant 10$	$\leqslant 12$
Cumulative frequency	4	9	16	37	45	48

  a Draw a cumulative frequency graph for the data.

  b Use your graph to estimate
    i the median
    ii the interquartile range of the times
    iii the 90th percentile.

  c Complete the grouped frequency table:

Time (min)	$0 < t \leqslant 2$	$0 < t \leqslant 4$				
Frequency						

  d Estimate the mean waiting time.

8 The number of customers visiting a shop is recorded over a period of 12 days:
26, 33, 28, 47, 52, 45, 93, 61, 37, 55, 57, 34
  a Find the median and the quartiles.
  b Determine whether there are any outliers.
  c Draw a box-and-whisker diagram for the data.

9 The heights of a group of 7 children were recorded to the nearest centimetre:
127, 119, 112, 123, 122, 126, 118
  a Find the mean and the variance of the heights.
  b Each child stands on a 35-centimetre-high stool. Find the mean and variance of their heights.

10 Theo is keeping a record of his travel expenses. The cost of each journey is $15 plus $3.45 per kilometre. The mean length of Theo's journeys is 11.6 km and the standard deviation of the lengths is 12.5 km. Find the mean and standard deviation of his cost per journey.

11 The frequency table summarizes data from a sample with mean 1.6. Find the value of x.

x	0	1	2	3
Frequency	5	6	8	x

12 A scientist measured a sample of 12 adult crabs found on a beach, measuring their shell length (*s*) and mass (*m*).

Shell length (cm)	Mass (g)	Shell length (cm)	Mass (g)
7.1	165	5.9	143
8.1	256	8.4	190
8.5	194	9.2	208
6.0	150	5.1	194
9.0	275	6.3	217
5.3	204	9.1	268

  a Using technology or otherwise plot a scatter diagram to illustrate the scientist's results.

  b The scientist later realized that the beach contains two species of crab – the Lesser European Crab and the Giant European Crab. Her research suggests that the Giant European crab tends to be heavier than similar-sized Lesser European Crabs. Find the equation for a regression line for the mass of the Giant European crab if its shell length is known.

  c Find the correlation coefficient for the data for the Giant European crab and comment on your result.

  d Estimate the mass of a Giant European Crab with shell length 8 cm.

  e Juvenile Giant European Crabs have a shell length of between 2 and 4 cm. Estimate the possible masses of these crabs and comment on the reliability of your results.

13 The one hour distances, in miles, covered by runners before (*x*) and after (*y*) going on a new training program are recorded.

The correlation between these two distances is found to be 0.84. The regression line is $y = 1.2x + 2$.

  a Describe the significance of

    i the intercept of the regression line being positive

    ii the gradient of the regression line being greater than 1.

  b If the previous mean distance is 8 miles, find the new mean.

Their trainer wants to have the data in km. To convert miles to km all the distances are multiplied by 1.6. The new variables in km are *X* and *Y*.

  c Find

    i the correlation between *X* and *Y*

    ii the regression line connecting *X* and *Y*.
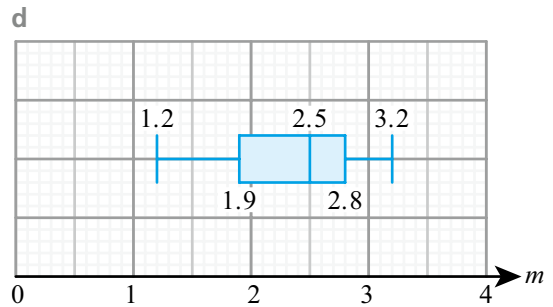
Chapter 6 Mixed Practice

1 a i systematic sampling

   ii e.g. Students may take books out on the same day each week.

 b i Each possible sample of 10 days has an equal chance of being selected.

   ii Representative of the population of all days.
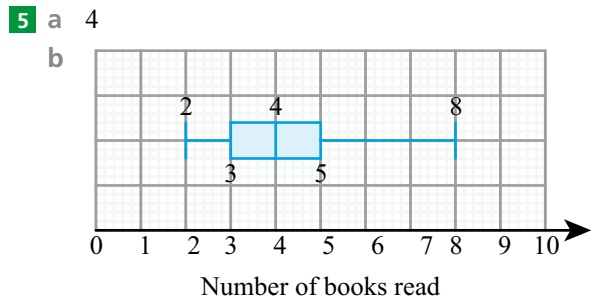
 c i 11

   ii 17.4

   iii 3.17

2 a
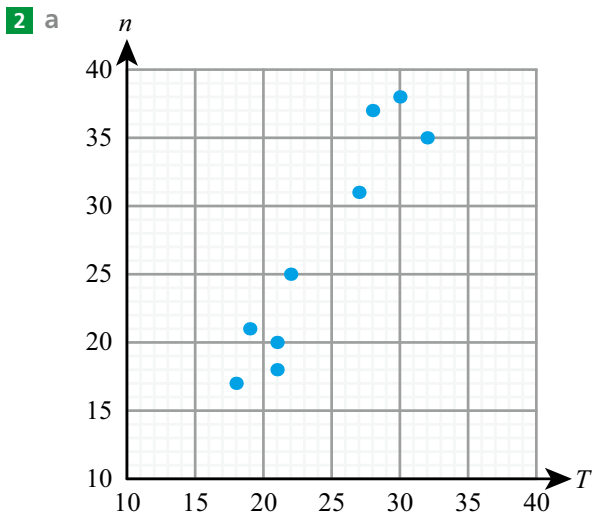
 b strong positive correlation
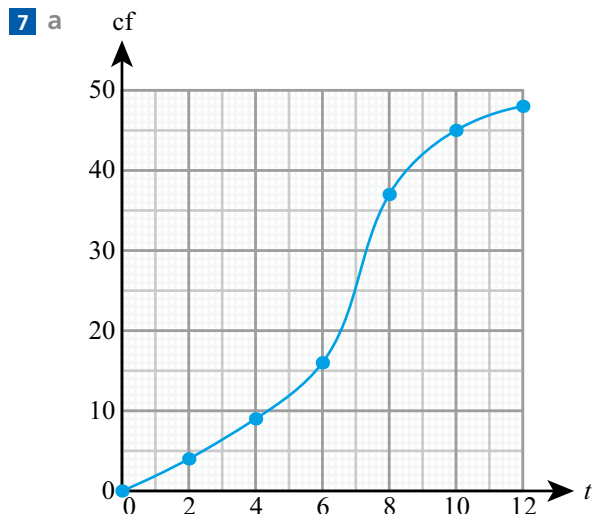
 c $n = 1.56T - 10.9$

 d 30.0

3 a 2.38 kg

 b cf

c 2.5 kg, 0.9 kg

d
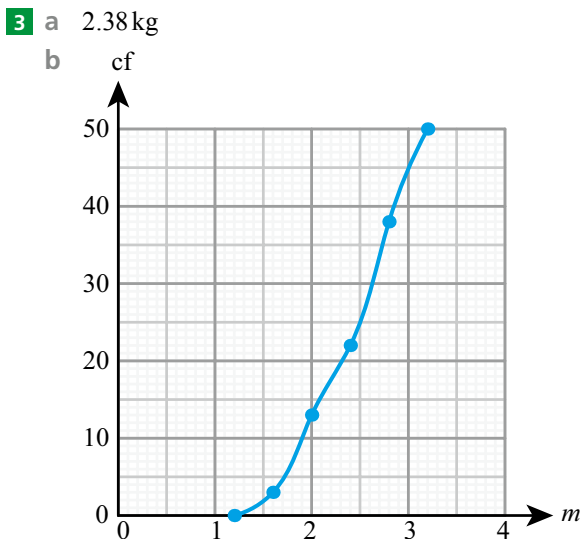
4 a discrete

 b 0

 c i 1.47

   ii 1.5

   iii 1.25

5 a 4

 b

Number of books read

 c 10

6 a 0.996

 b $a = 3.15,\ b = -15.4$

 c 66.5

7 a cf

 b i 6.9 minutes

   ii 2.9 minutes

   iii 9.3 minutes

c

Time	Freq
$0 \leqslant t < 2$	4
$2 \leqslant t < 4$	5
$4 \leqslant t < 6$	7
$6 \leqslant t < 8$	11
$8 \leqslant t < 10$	8
$10 \leqslant t < 12$	3

d 6.21 minutes

8 a med = 46, $Q_1 = 33.5$, $Q_3 = 56$

b yes (93)

c
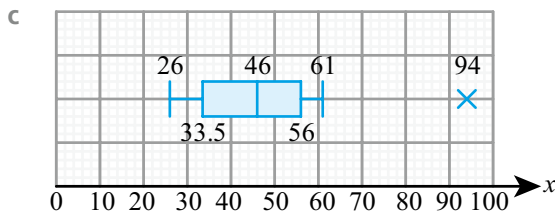
9 a 121 cm, 22.9 cm²

b 156 cm, 22.9 cm²

10 mean = $55.02, sd = $43.13

11 $x = 6$

12 a
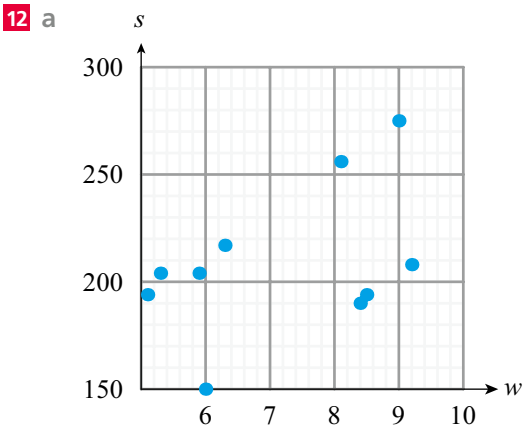
b $w = 19.1s + 99.0$

c 0.994 – strong positive correlation

d 252 g

e 137 g to 175 g; extrapolating from the data so not reliable.

13 a i Athletes generally do better after the programme.
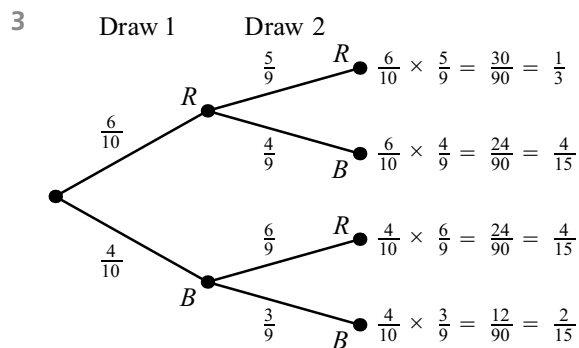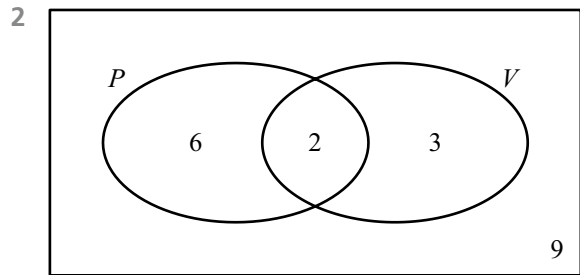
ii Better athletes improve more.

b 11.6 miles

c i 0.84

ii $Y = 1.2X + 3.2$

Chapter 7 Prior Knowledge

1 a $\{2, 3, 4, 5, 6, 7, 8, 9\}$

b $\{5, 8\}$

2

3

Exercise 7A

1 a $\dfrac{3}{10}$ **b** $\dfrac{2}{15}$

2 a $\dfrac{1}{5}$ **b** $\dfrac{3}{5}$

3 a $\dfrac{1}{2}$ **b** $\dfrac{1}{3}$

4 a $\dfrac{1}{4}$ **b** $\dfrac{1}{13}$

5 a $\dfrac{1}{26}$ **b** $\dfrac{3}{26}$

6 a 0.94 **b** 0.55

7 a $\dfrac{11}{20}$ **b** $\dfrac{23}{40}$

8 a 0.85 **b** 0.13

9 a $\dfrac{47}{120}$ **b** $\dfrac{41}{48}$

10 a 0.73 **b** 0.66

11 a 0.44 **b** 0.11

12 a 4 **b** 27

13 a 12 **b** 6

14 a 4.8 **b** 7.5