

Statistics GCSE

Paper 2

Edexcel Foundation - 2026

Foundation Tier

Variant 1 (same as video)

1ST0/2F

# Answers

Visit our website for tutorials on each question.

[www.statsgcse.com](http://www.statsgcse.com)

1



Ethan has 6 number tiles shown above.

All of the tiles are placed inside a bag then a random tile is drawn from the bag.

- (a) Select the word describes the likelihood that the tile has a 2 on it.

(1 mark)

Select **one** box.

- impossible
- unlikely
- likely
- evens

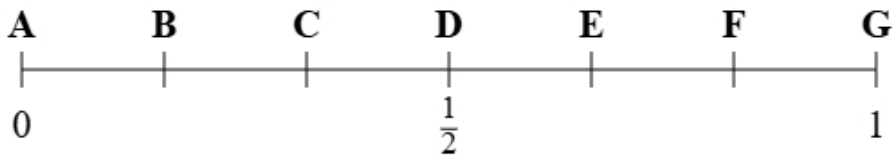
- (b) Select the word describes the likelihood that the tile has a 4 on it.

(1 mark)

Select **one** box.

- likely
- unlikely
- impossible
- evens

(c)



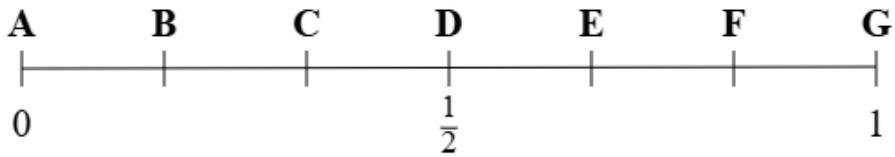
Using the probability scale, write down the letter that shows probability that the tile has a 1 on it.

(1 mark)

Select **one** box.

- B
- C
- A
- D

(d)



Using the probability scale, write down the letter that shows probability that the tile has a 2 or a 3 on it.

(1 mark)

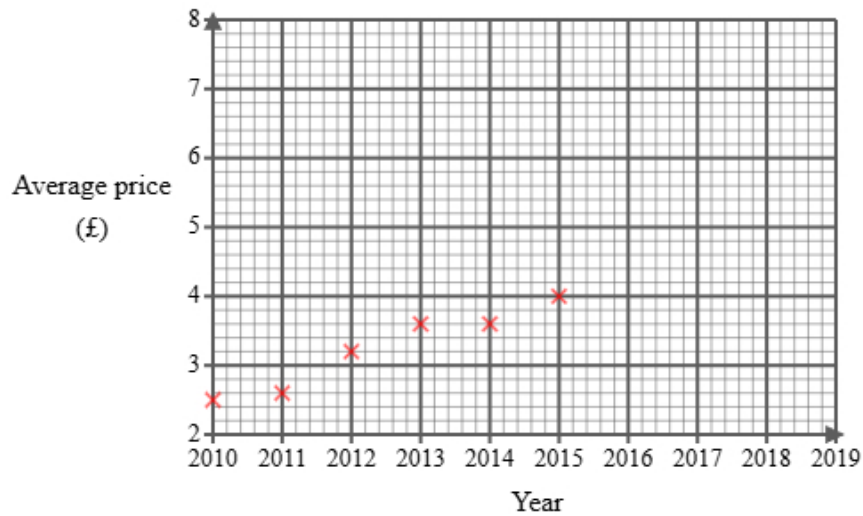
Select **one** box.

- F
- D
- G
- E

2 Omar found the following information about the average price of a single bus fare in England.

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Average price (£)	2.50	2.60	3.20	3.60	3.60	4.00		4.80	5.00	5.50

He did not find the price for 2016 and has started to draw a graph for the data.

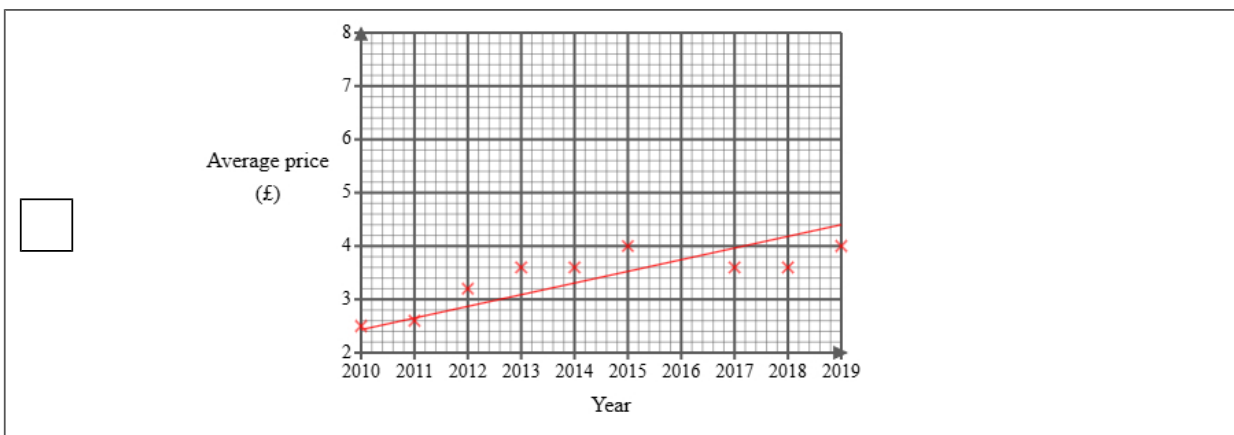
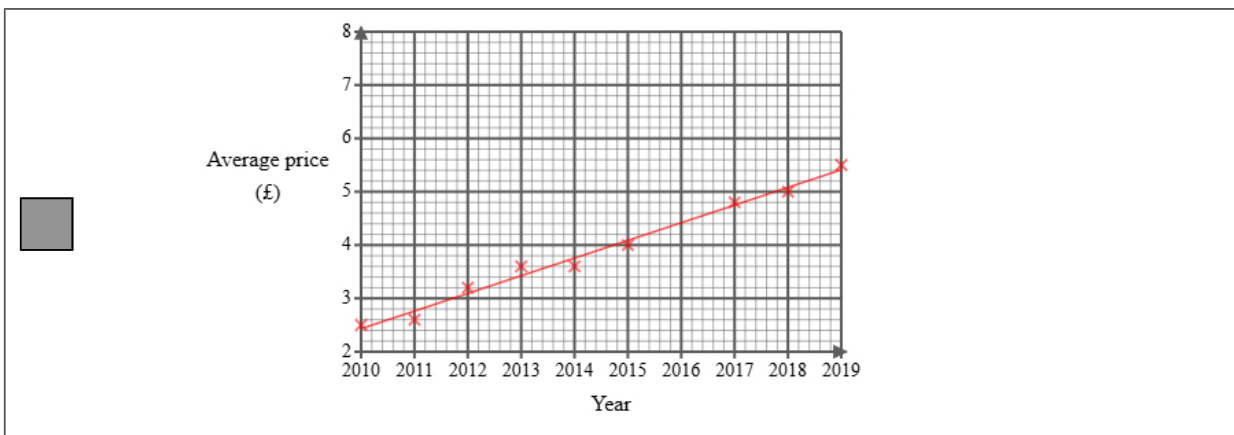
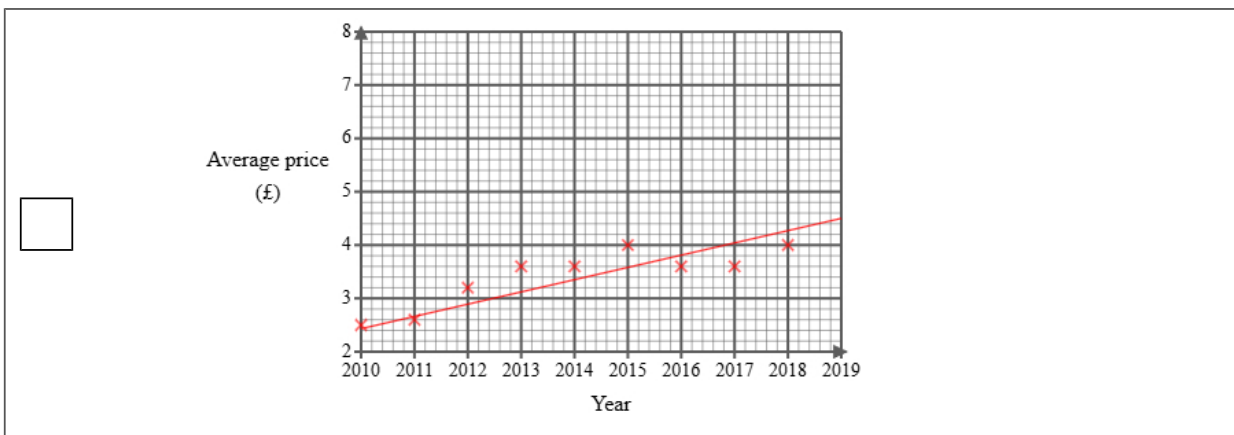
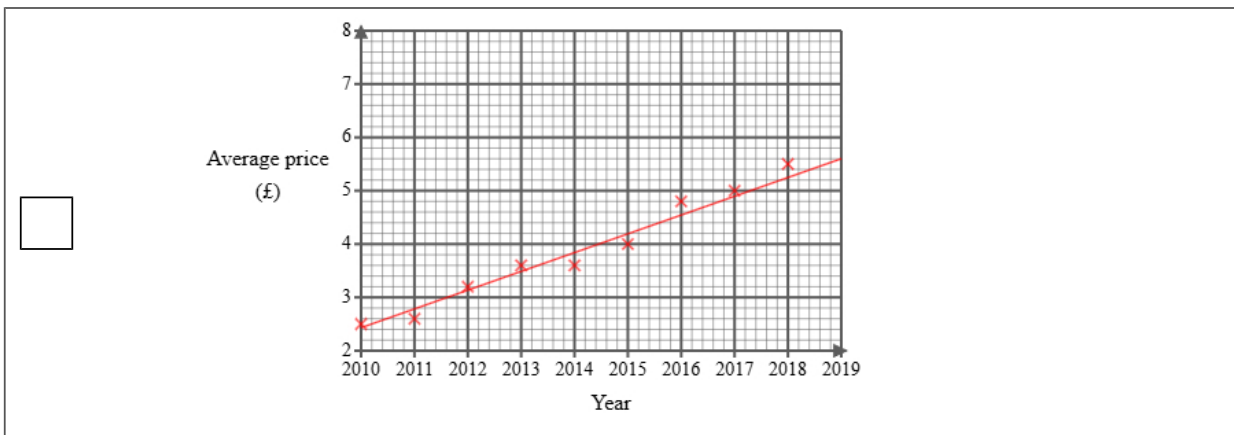


Omar then used statistical software to calculate the equation for the trend line.

- (a) (i) Plot the average price for each of 2017, 2018 and 2019  
(ii) Draw a trend line for Omar's data  
(iii) Describe the trend in the average price of a single bus fare in the UK from 2010 to 2019

(4 marks)

Select the correct answer.



Select the correct boxes.

- Decreasing
- Negative correlation
- Increasing
- Positive correlation

(b) The gradient of Omar's trend line is 0.33

Interpret this gradient.

(1 mark)

Select **one** box.

- The average price decreases.
- The average price started at 33 pence in 2010.
- The average price increases.
- The average price increases per year.

(c) Explain whether or not the scale used on the 'average price' axis could make the graph misleading.

(2 marks)

Select the **two** correct statements (**two** statements are incorrect).

- The graph could be misleading.
- The 'average price' axis does not start from zero.
- The graph is not misleading.
- The graph is labelled correctly.

(d) Omar draws the trend line onto the graph.

He suggests that the trend line could be used to estimate the average price for 2016 **and** 2020

Explain whether each of these estimates would be reliable.




You should **not** work them out.

(3 marks)

Select the **two** correct statements (**two** statements are incorrect).

- 2020 would not be reliable because it is outside the range of the data.
- 2020 would be reliable because it is within the data.
- 2016 would not be reliable because it is outside the range of the data.
- 2016 would be reliable because it is within the data.

- 3 A bakery tracks the number of loaves of bread sold from Monday to Thursday. This helps them understand which days are busier and plan their stock accordingly.

Monday	
Tuesday	
Wednesday	
Thursday	
Friday	

**Key:**

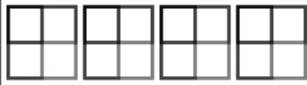





represents 8 loaves of bread

- (a) On Thursday, the number of loaves of bread sold was 32.  
Show this information on the pictogram.

(1 mark)

Select the correct answer.

<input checked="" type="checkbox"/>	Thursday	
<input type="checkbox"/>	Thursday	
<input type="checkbox"/>	Thursday	
<input type="checkbox"/>	Thursday	

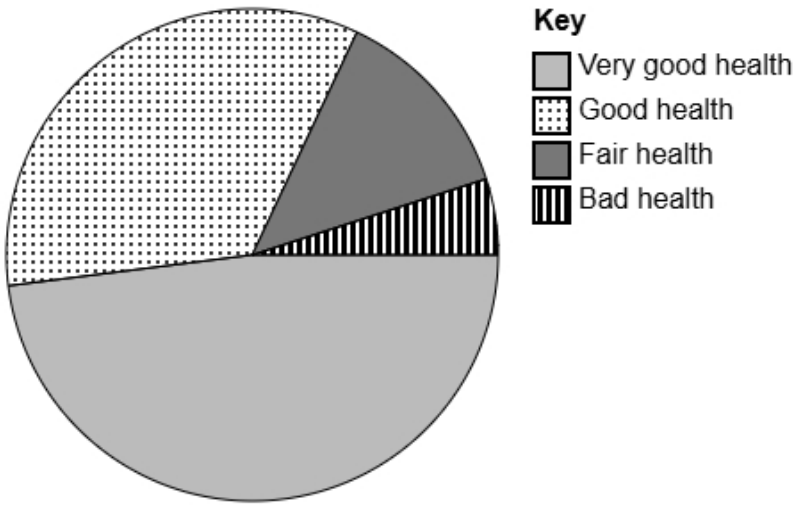
- (b) Felix suggests redrawing the pictogram using a key with a whole-square representing 5 loaves of bread. Explain why this key would **not** be suitable.

(1 mark)

Select **one** box.

- Tuesday shows 6 loaves of bread. This would be very difficult to show because 6 has a remainder 1 when divided by 5.
- If Felix uses 5 loaves of bread for the key this will lead to less sales being shown.
- The key must always be an even number.
- This would be much better because you can fit more squares on.

4 The accurately drawn pie chart shows information about how people in the UK rated their health in 2021.



(a) Explain how you can tell that most people viewed themselves as 'Very good health' in the UK in 2021 using the pie chart.

(1 mark)

Select **one** box.

- 'Very good health' is at the bottom of the pie chart.
- 'Very good health' is the first value in the key.
- 'Very good health' is the most positive response.
- 'Very good health' has the largest sector.

(b) The population in the UK in 2021 was estimated to be 67 million.

Calculate an estimate for the number of people in the UK in 2021 who rate themselves as having 'Good health'.

Round your answer to the nearest million.

(2 marks)

\_\_\_\_\_ 23 million

5 Sarah owns a cinema.

She wants to collect information about types of movies liked by people in her city.

The following list gives the information she is going to collect about people's favourite movies:

genre

average movie length

minimum age rating

(a) Select the information that is categorical data from the list.

(1 mark)

Select **one** box.

average movie length

minimum age rating

genre

(b) Select the information that is discrete data from the list.

(1 mark)

Select **one** box.

genre

minimum age rating

average movie length

(c) Sarah would like to send a questionnaire to 40 of her customers.

She has a list of all 200 of her customers.

Explain how Sarah can select a systematic sample of 40 people from her list of customers.

(2 marks)

Select the **two** correct statements (**two** statements are incorrect).

Calculate a random number between 1 and 4 to use a starting point.

Select every 5th person.

Calculate a random number between 1 and 5 to use a starting point.

Select every 4th person.

6 Mia is a student and wants to study after-school hobbies.

She would like to find out the most common after-school hobby in her school.

Mia decides to do convenience sampling outside the school gate an hour after school and uses the data collection sheet shown here:

Hobby	Tally
Reading	
Gaming	
Sport	
Music	

(a) State the population for this study.

(1 mark)

Select **one** box.

- A selection of students in Mia's school
- The students she asks
- All the students in Mia's school
- All students in the UK

(b) Describe what is meant by a convenience sample.

(1 mark)

Select **one** box.

- Sampling people in proportion to a characteristic of the population
- Sampling every  $n$ th person from a list
- Sampling people randomly
- Sampling people who are easiest to reach

(c) Give **one** disadvantage of convenience sampling.

(1 mark)

Select **one** box.

- It is not representative
- It is expensive to do
- It takes too much time
- It requires a large population

(d) Discuss whether this data collection sheet is appropriate for Mia to collect the data.

(2 marks)

Select the **two** correct statements (**two** statements are incorrect).

- Some students may not have an after-school hobby.
- Some students may not play sport after school.
- It will make collecting data very difficult.
- It will make it easy for Mia to analyse the data.

(e) After collecting the data, Mia would like to display the data in a diagram.

Discuss whether or not a stem and leaf diagram would be suitable.

(2 marks)

Select the **two** correct statements (**two** statements are incorrect).

- because the data is quantitative.
- because the data is qualitative.
- It is suitable
- It is not suitable

7 An instructor wants to get feedback on a webinar on digital safety they ran last week.

72 people attended the webinar.

The instructor plans to give a questionnaire to a sample of 15 of the people who attended the webinar.

One of the questions on the questionnaire is:

To what extent do you agree with the statement, the instructor provided useful information?

Use a scale from 0 to 5, where 0 means **strongly disagree** and 5 means **strongly agree**.

(a) Select the word from the list that best describes the type of data collected by this question.

(1 mark)

Select **one** box.

bivariate

continuous

ordinal

(b) Explain how the instructor could use a list of random numbers to choose a simple random sample of 15 people who attended the webinar.

(3 marks)

Select the **three** correct statements (**three** statements are incorrect).

List the people in alphabetical order based on their surname.

Select the people who were assigned the highest number.

Select people with the corresponding number from random number list.

Assign a number to all the people.

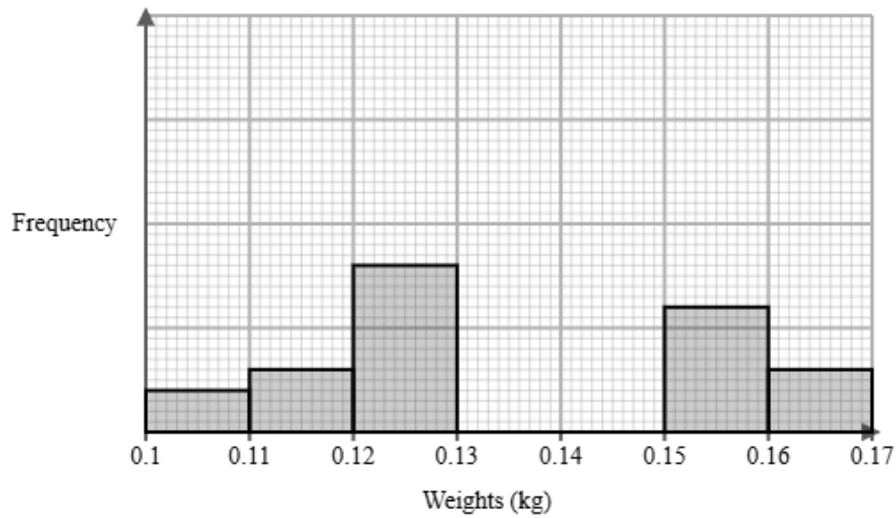
Select 15 numbers ignoring any numbers that have repeated or are out of range.

Select 15 people from a hat.

8 An orchard manager measured the weights of red apples and green apples in region A.

They recorded the weights after 6 months.

The incomplete histogram and grouped frequency table give information about the weights of red apples in region A.



Weights $w$ (kg)	Frequency
$0.10 < w \leq 0.11$	2
$0.11 < w \leq 0.12$	3
$0.12 < w \leq 0.13$	8
$0.13 < w \leq 0.14$	14
$0.14 < w \leq 0.15$	9
$0.15 < w \leq 0.16$	
$0.16 < w \leq 0.17$	

(a) Use the information in the histogram to complete the table.

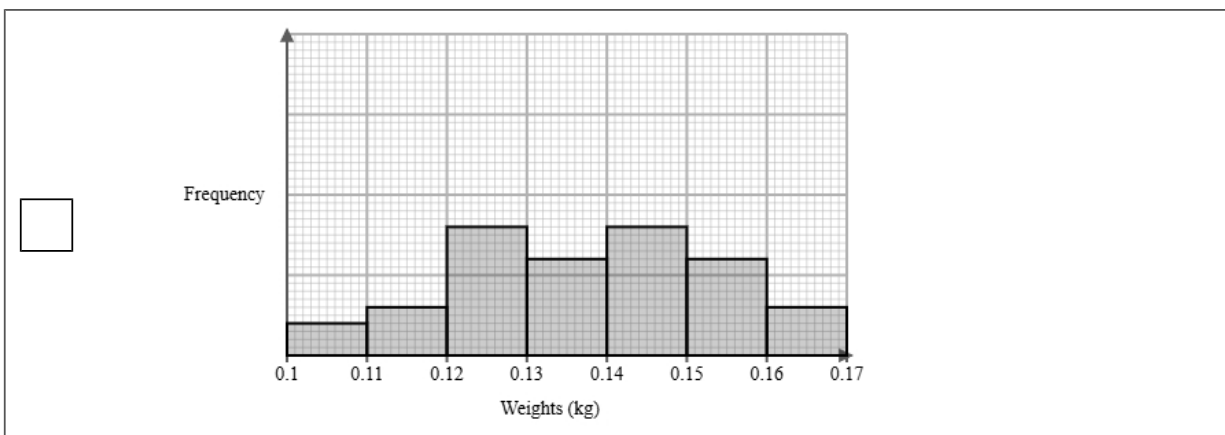
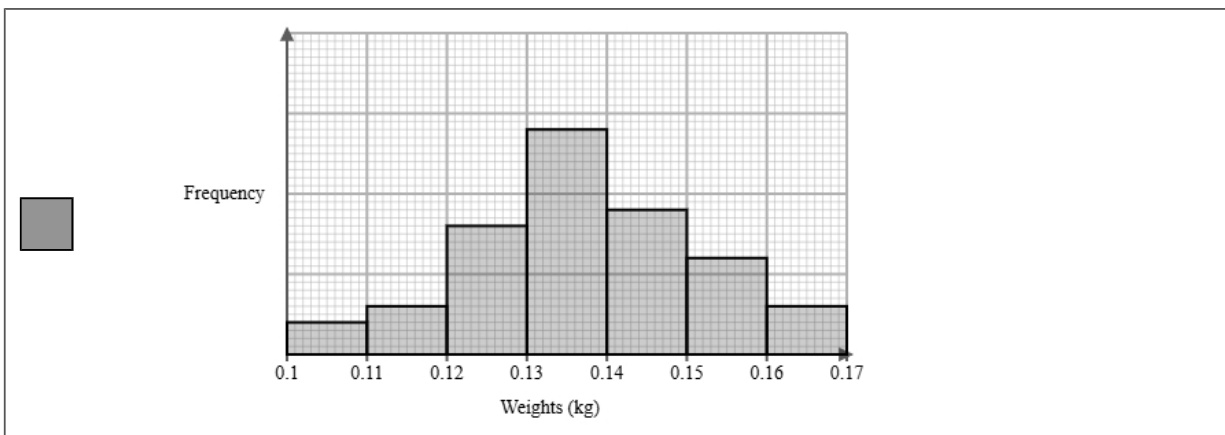
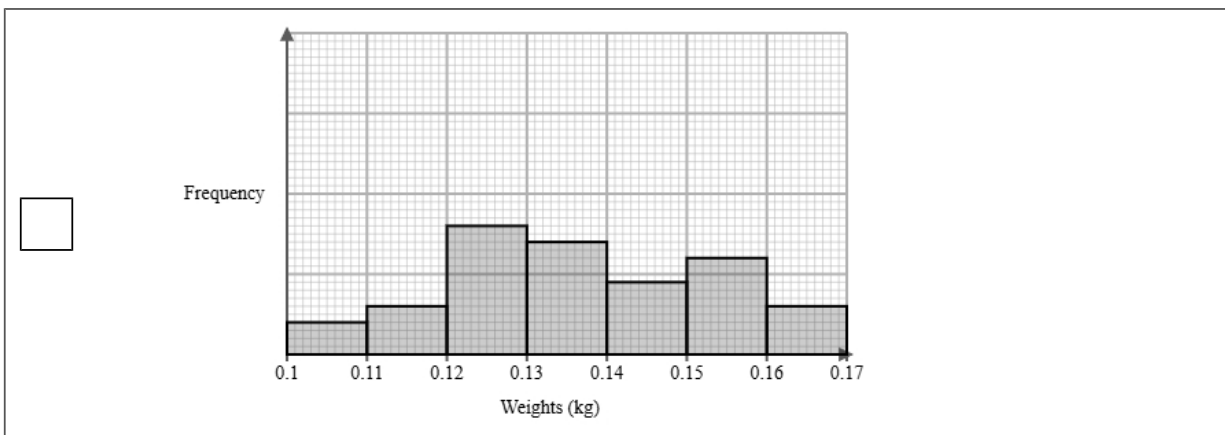
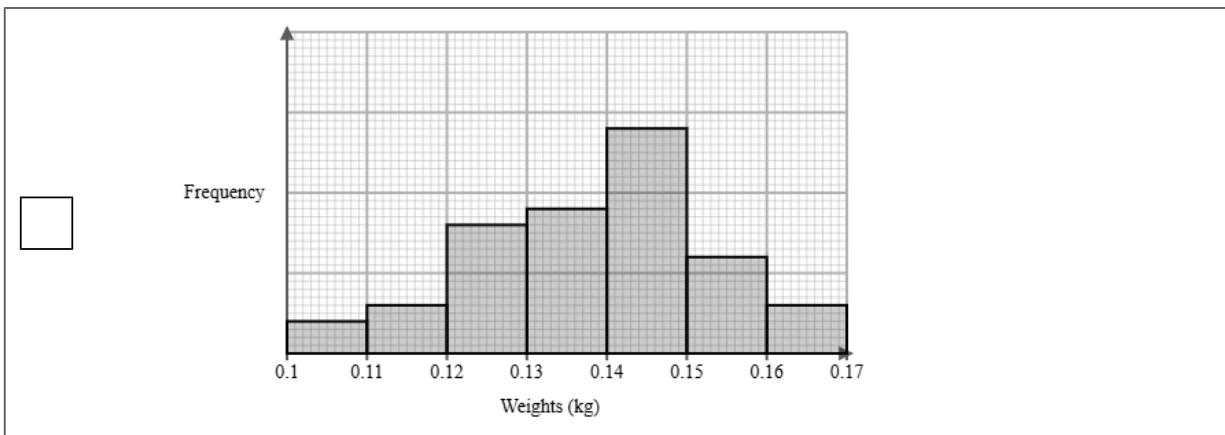
(2 marks)

Weights $w$ (kg)	Frequency
$0.10 < w \leq 0.11$	2
$0.11 < w \leq 0.12$	3
$0.12 < w \leq 0.13$	8
$0.13 < w \leq 0.14$	14
$0.14 < w \leq 0.15$	9
$0.15 < w \leq 0.16$	<u>6</u>
$0.16 < w \leq 0.17$	<u>3</u>

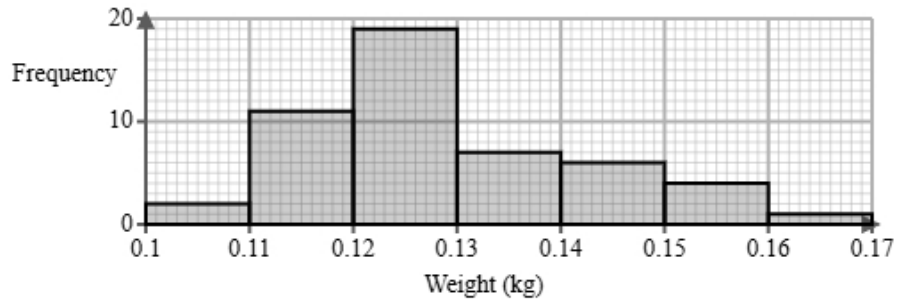
(b) Use the information in the table to complete the histogram.

(2 marks)

Select the correct answer.



(c) The histogram below shows data on the weights of green apples after 6 months.



Identify and interpret the type of skew shown in the histogram for green apples.

(2 marks)

Select the **two** correct statements (**two** statements are incorrect).

- The mean weight is smaller than the median.
- Negative skew.
- The mean weight is larger than the median.
- Positive skew.

- (d) The orchard manager also measured the weights of red apples and green apples in region B. The grouped frequency table below gives information about the weights of red apples and green apples in region B.

Weights $w$ (grams)	Frequency	
	red apples	green apples
$80 < w \leq 100$	9	9
$100 < w \leq 120$	13	19
$120 < w \leq 140$	12	8
$140 < w \leq 160$	3	2
Total	37	38

The estimate of the mean for red apples is calculated to be 114.9 g to 1 decimal place.

Susan uses the estimate of the means for red apples and green apples to conclude that the red apples weigh less than green apples.

Discuss whether or not Susan is correct and give **one** limitation of your conclusion.

You must show your working.

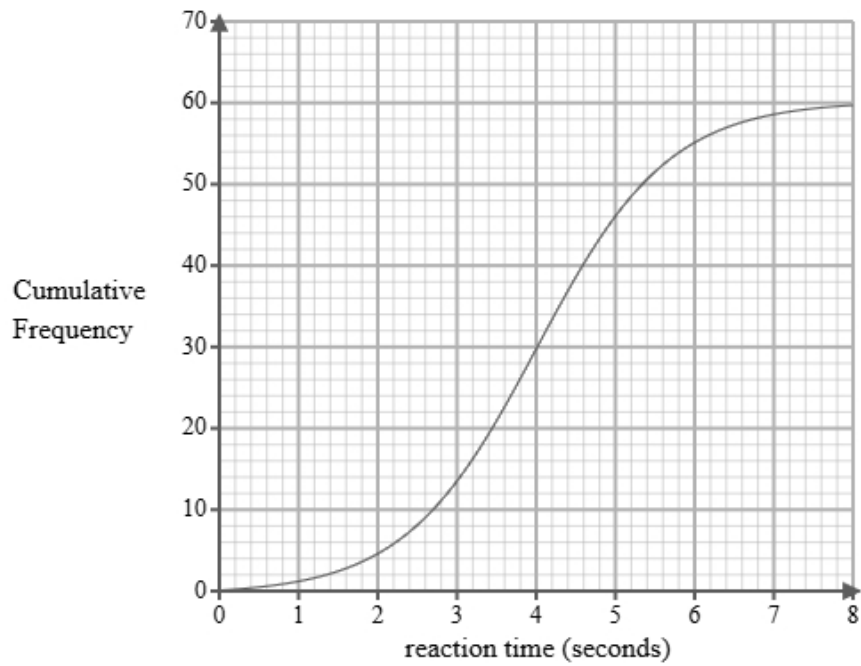
(4 marks)

Mean of green apples = 111.6

Select the **two** correct statements (**two** statements are incorrect).

- Susan is incorrect.
- Susan is correct.
- We cannot be sure because both means are in the same class interval.
- These are only estimates.

- 9 A researcher measures the reaction times, in seconds, of 60 students completing a computer task. A cumulative frequency diagram is drawn from the data.



Complete the table below from the cumulative frequency diagram.

Lower quartile	Median	Upper quartile

(2 marks)

Select the correct answer.

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	2.1	4.6	6.5

<input checked="" type="checkbox"/>	Lower quartile	Median	Upper quartile
	3.1	4	4.9

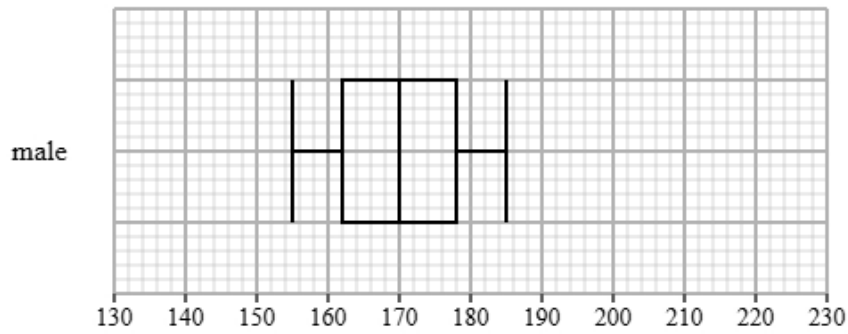
<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	2.6	4.6	5.6

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	3.9	4	4.5

10 Sophie recorded the heights of male and female students in a school.

Both groups were measured using the same method.

The box plot shows information about the heights for the male students.



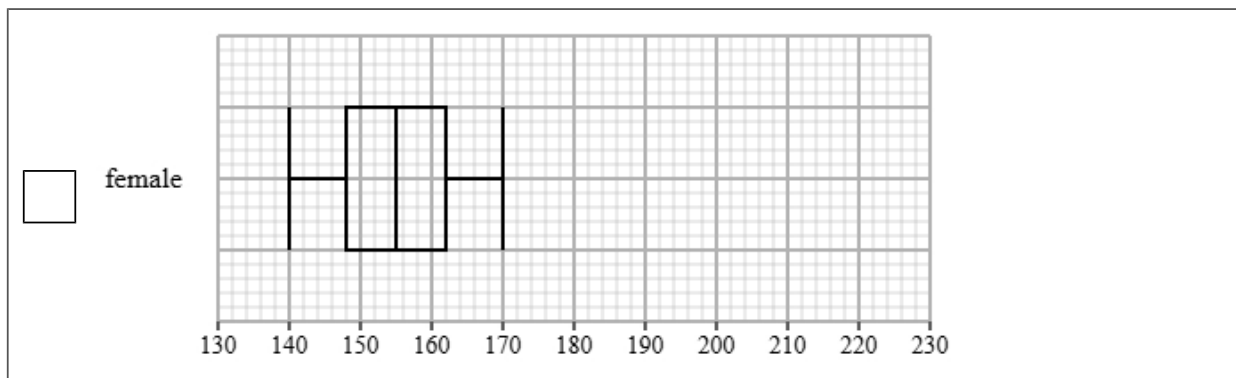
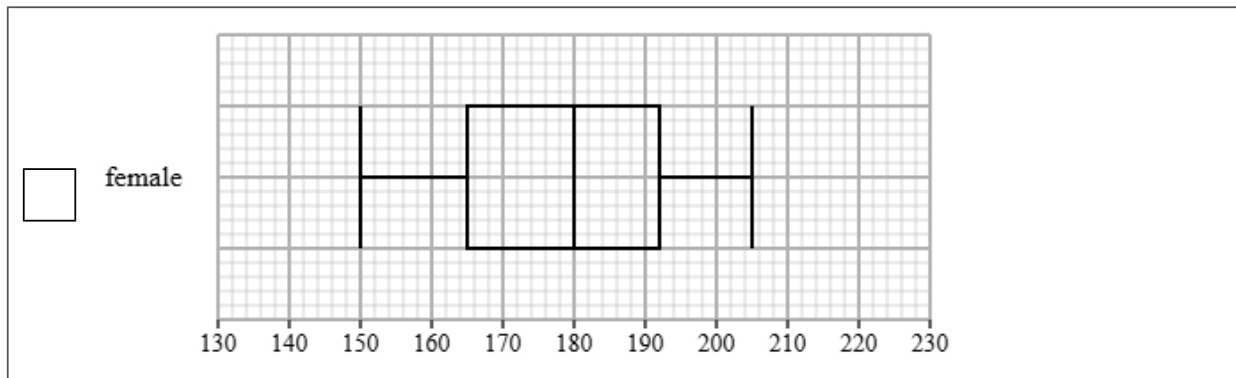
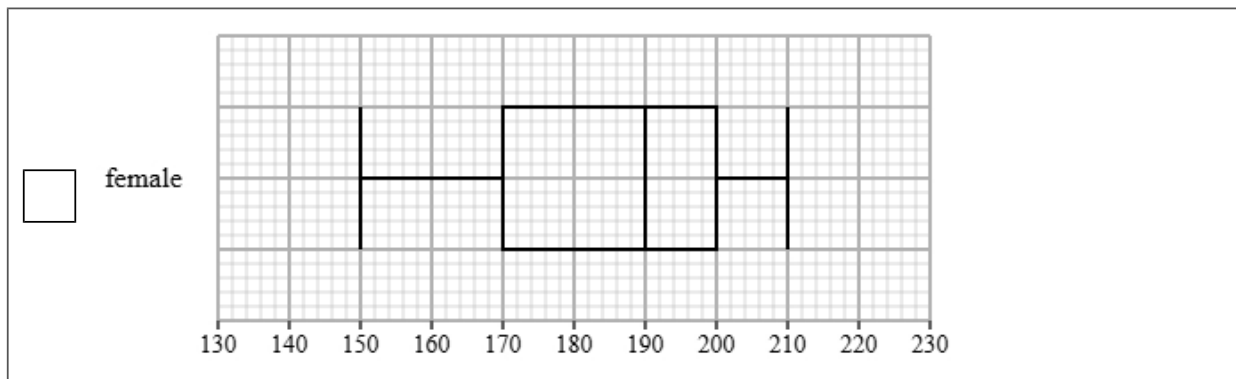
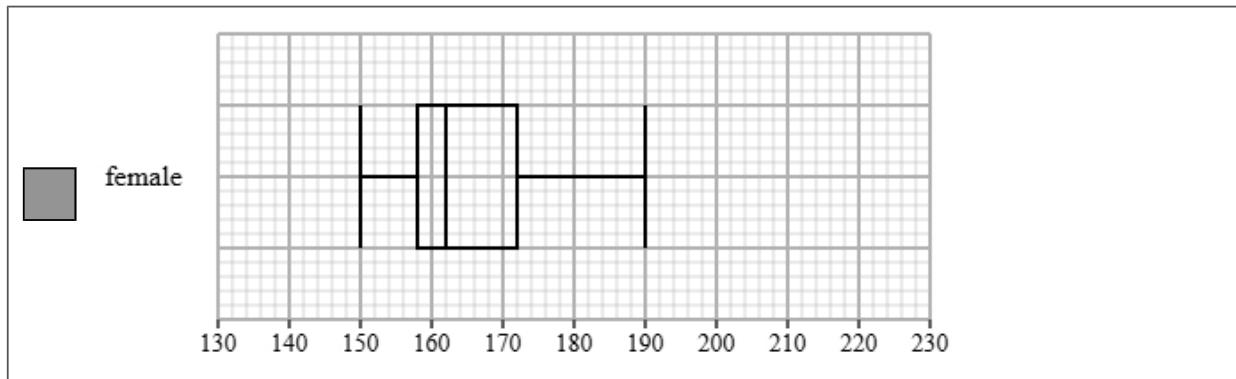
The table gives information about the heights for the female students.

Least tall	Lower quartile	Median	Upper quartile	Most tall
150	158	162	172	190

(a) Draw a box plot for the heights for the female students.

(2 marks)

Select the correct answer.



(b) Compare the two distributions of heights.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median heights for male and female students are equal.
- The median heights for male students is greater than female students.
- The median heights for male students is lower than female students.

Select **one** box.

- The IQR is bigger.
- The IQR for the heights of the male and female students are equal.
- The IQR for the heights of the male students is greater than female students.
- The IQR for the heights of the male students is lower than female students.

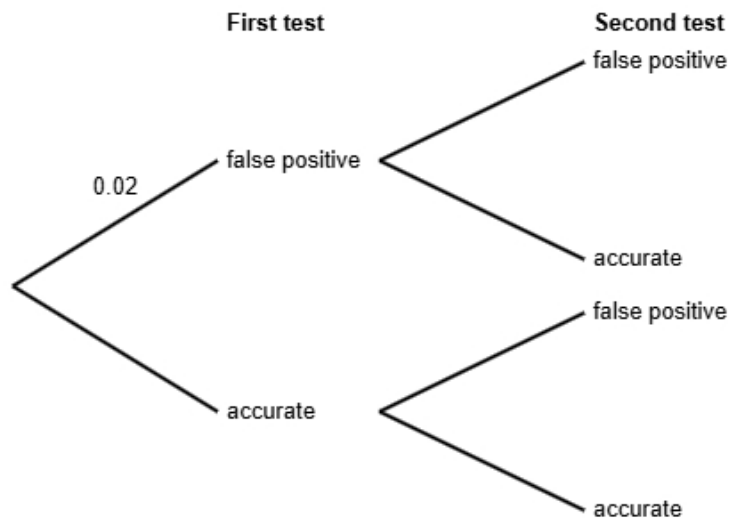
Select **one** box.

- The skews for the heights of the male and female students are both positive.
- The skew for the heights of the male students is symmetrical and the skew for the female students is negative.
- The skew for the heights of the male students is symmetrical and the skew for the female students is positive.
- The skews for the heights of the male and female students are both symmetrical.

Select **one** box.

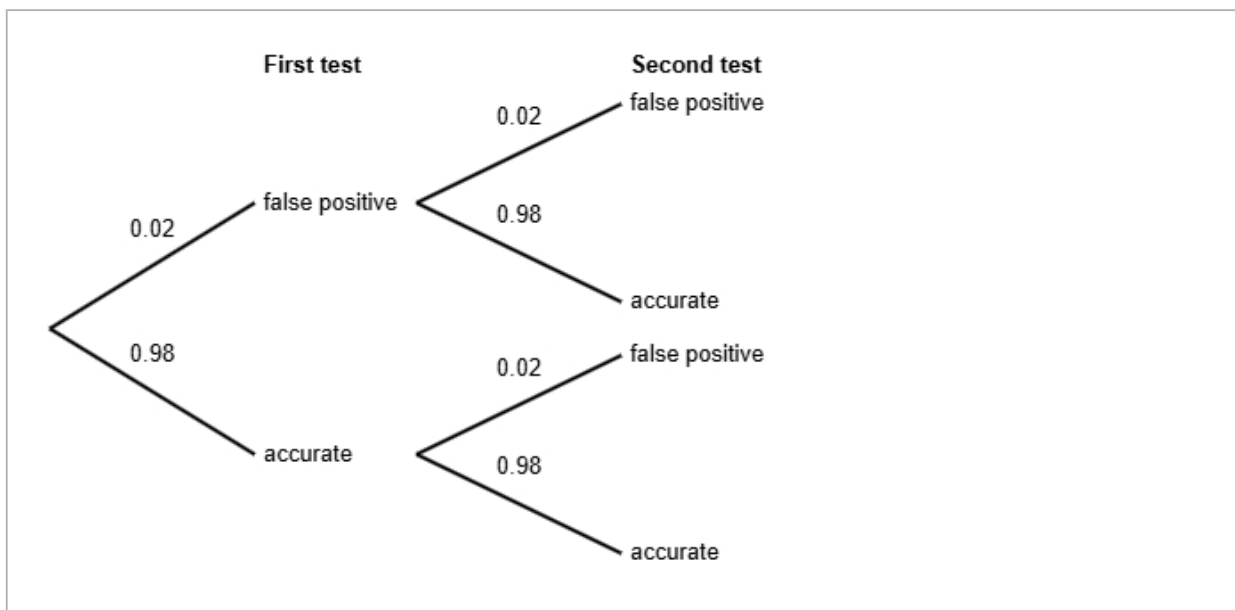
- The heights for the male students are less spread out than the female students.
- The male students are more skewed than female students.
- The male students are on average taller than the female students.
- The male students are on average shorter than the female students.

- 11** It is estimated that 2% of a certain type of medical test gives a false positive result.  
 The remaining tests provide accurate results.  
 David took two such tests.  
 He does not know if each test result is false positive or accurate.



- (a) Complete the probability tree diagram.

(2 marks)



(b) Find the probability that both of David's test results are accurate.

(2 marks)

0.9604

(c) David states that the probability that exactly one of the tests is false positive is less than 4%. Find out whether or not David is correct.

(3 marks)

$$\begin{aligned} P(\text{false positive AND accurate}) &= 0.02 \times 0.98 \\ &= 0.0196 \\ P(\text{exactly one of the tests is false positive}) &= 0.0196 \times 2 \\ &= 0.0392 \\ 0.0392 &= 3.92\% \end{aligned}$$

Select **one** box.

- The probability that exactly one of the tests is false positive is more than 4%, so David is not correct.
- The probability that exactly one of the tests is false positive is less than 4%, so David is not correct.
- The probability that exactly one of the tests is false positive is more than 4%, so David is correct.
- The probability that exactly one of the tests is false positive is less than 4%, so David is correct.

12 The table shows information about cars for sale in Cambridge.

number of doors	number of cars
2	340
3	180
4	260
5	520
6 or more	500
Total	1800

A researcher wants to investigate the price of these cars and takes a stratified sample of 90 cars according to the number of doors.

(a) The researcher says the mode of the number of doors for these cars is 5.

Explain how the researcher knows this.

(1 mark)

Select **one** box.

5 is the middle number.

5 doors has the highest frequency.

5 is the difference between the largest and smallest number.

5 cars has the highest frequency.

(b) Work out the number of cars in the sample for each number of doors.

number of doors	number of cars in the sample
2	
3	
4	
5	
6 or more	

(3 marks)

number of doors	number of cars in the sample
2	17
3	9
4	13
5	26
6 or more	25

(c) Describe how the 90 cars in the sample should be selected.

(3 marks)

Select the **three** correct statements (**three** statements are incorrect).

- Ensure that all 1800 cars are included in the sample.
- Number each of the cars, and then use the random numbers to select the required amount of cars.
- Select the first 90 cars.
- Use a sampling frame for each strata.
- Generate random numbers, remove repeats or numbers out of range.
- Complete two of the strata.

- 13** Aisha works for an environmental agency. She has been tasked with investigating air pollution levels near schools.

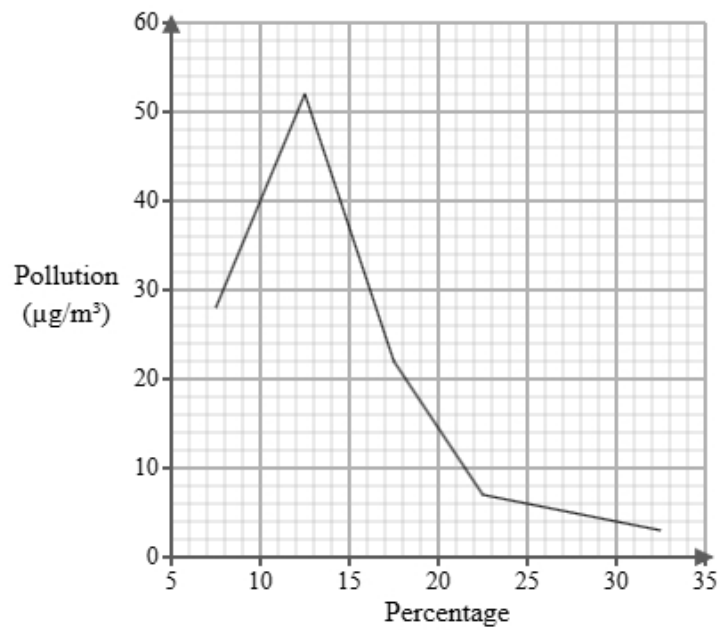
Below is a section of the spreadsheet she used to record her findings.

Pollution ( $\mu\text{g}/\text{m}^3$ )	Percentage
$5 < p \leq 10$	6
$10 < p \leq 15$	eight
$15 < p \leq 20$	8
$20 < p \leq 25$	117
$25 < p \leq 30$	48
$30 < p \leq 35$	13
Total	100

Aisha cleans the data to create the table below.

Pollution ( $\mu\text{g}/\text{m}^3$ )	Percentage
$5 < p \leq 10$	6
$10 < p \leq 15$	8
$15 < p \leq 20$	8
$20 < p \leq 25$	17
$25 < p \leq 30$	48
$30 < p \leq 35$	13
Total	100

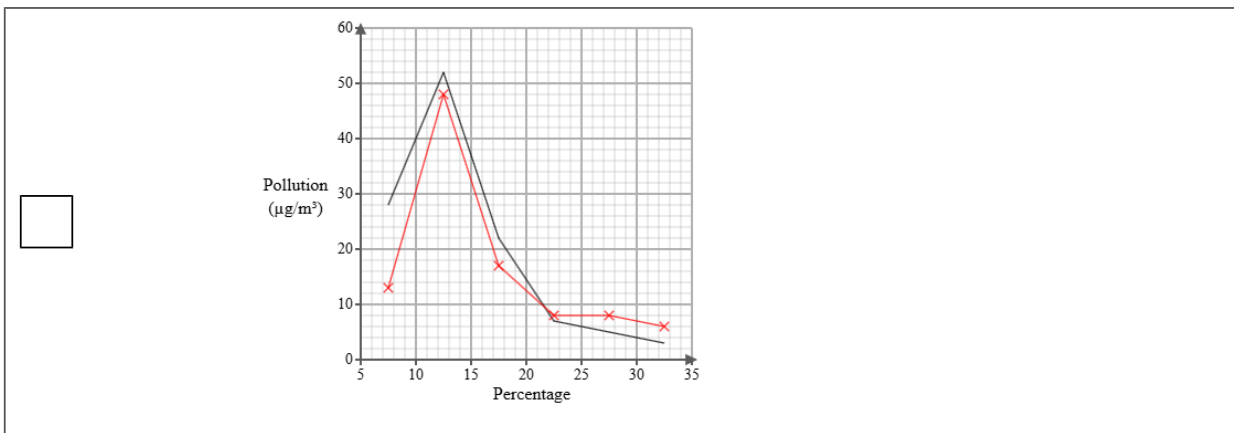
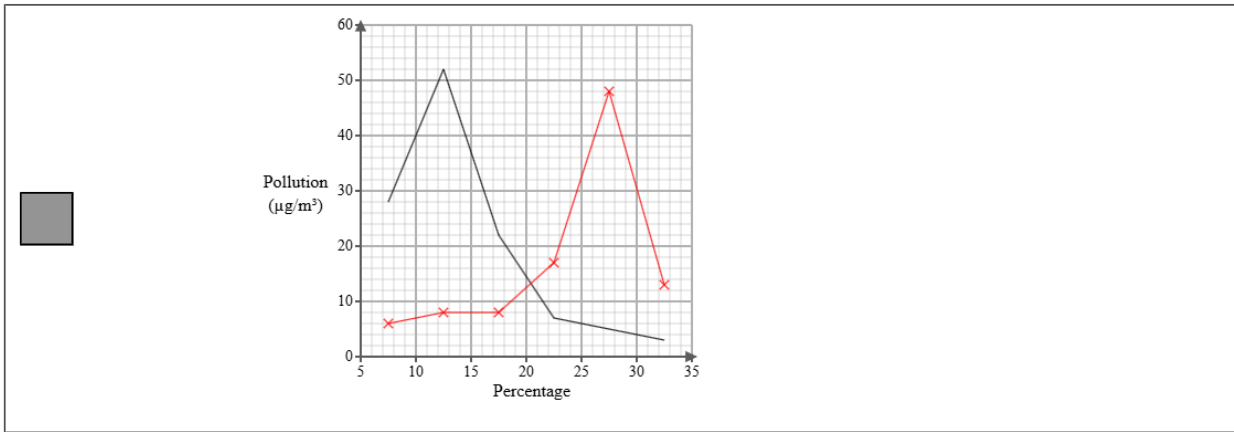
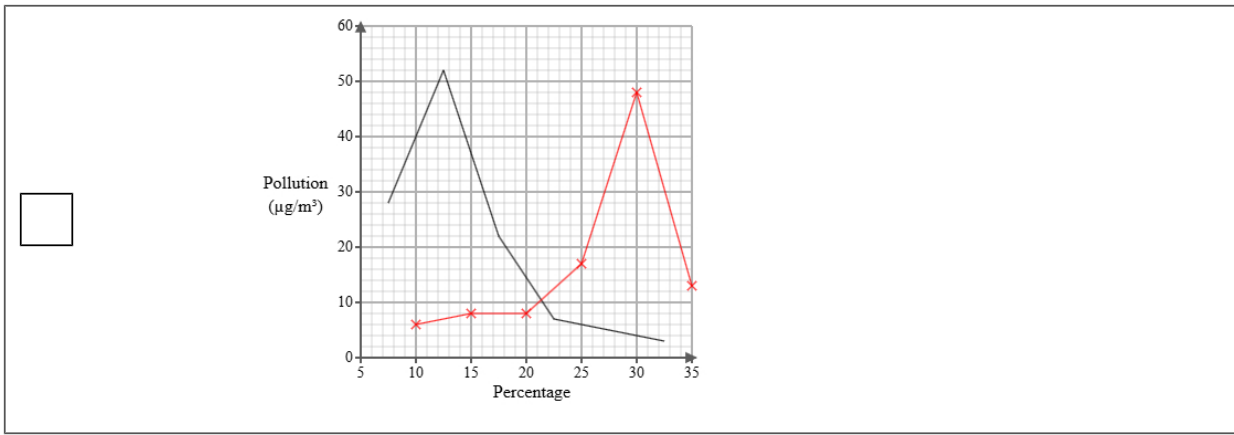
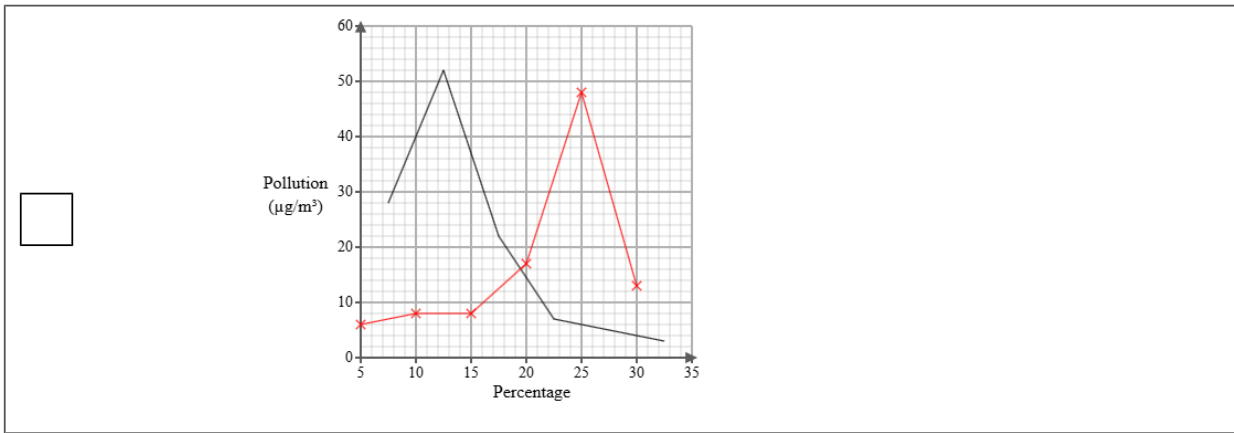
A frequency polygon has been drawn for air pollution levels near parks.



- i) On the same graph, draw the frequency polygon for air pollution levels near schools.
- ii) Using the two frequency polygons, compare the skew of the distributions and explain what your comparison means in context.

(4 marks)

Select the correct answer.



Select the **two** correct statements (**four** statements are incorrect).

- This means that for the parks the pollution is mainly at the upper end of the distribution and for the schools the pollution is mainly at the lower end of the distribution.
- This means that for the parks the pollution is mainly at the lower end of the distribution and for the schools the pollution is mainly at the upper end of the distribution.
- The distribution of air pollution levels near parks is symmetrical whereas the distribution of air pollution levels near schools is positively skewed.
- The distribution of air pollution levels near parks is positively skewed whereas the distribution of air pollution levels near schools is negatively skewed.
- This means that for the parks the pollution is equally spread out on either side of the median and for the schools the pollution is mainly at the upper end of the distribution.
- The distribution of air pollution levels near parks is negatively skewed whereas the distribution of air pollution levels near schools is positively skewed.

- 14 The table shows information about the consumer price index (CPI) and TV Licence price (£) in the United Kingdom for Jan 2000, Jan 2010 and Jan 2020.

	Jan 2000	Jan 2010	Jan 2020
consumer price index	100	123	153
TV Licence price (£)	104	145.5	157.5

Describe how the increase in TV Licence price (£) compares with the CPI over the ten years to Jan 2010 and over the twenty years to Jan 2020.

(5 marks)

Select the **four** correct statements (**four** statements are incorrect).

$\frac{145.5}{123} \times 100 = 118$  (nearest integer)

$\frac{157.5}{153} \times 100 = 105$  (nearest integer)

Between Jan 2000 and Jan 2010 the change in price was less than the CPI.

Between Jan 2000 and Jan 2020 the change in price was more than the CPI.

Between Jan 2000 and Jan 2010 the change in price was more than the CPI.

Between Jan 2000 and Jan 2020 the change in price was less than the CPI.

$\frac{157.5}{104} \times 100 = 151$  (nearest integer)

$\frac{145.5}{104} \times 100 = 140$  (nearest integer)

**15** Sarah is investigating how the age in years,  $x$ , affects the resale price (£),  $y$  for two types of laptops, model A and model B.

She found ten laptops of each type and recorded their age and resale price and plotted each on scatter diagrams.

She then drew a line of best fit on each diagram and found the gradient and y-intercept of each line.

Here are the results:

Model	Gradient of line of best fit	y-intercept of line of best fit
A	-120	1500
B	-95	2000

Interpret and compare these results in context.

(5 marks)

Select the **five** correct statements (**five** statements are incorrect).

- Model A reduces in resale price more per year than Model B.
- Model A has a greater initial resale price.
- Model A changes in resale price by £1500 per year.
- Model B changes in resale price by £2000 per year.
- Model A reduces in resale price less than Model B.
- Both laptops increase in resale price as the age increase.
- Model A reduces in resale price by £120 per year.
- Both laptops decrease in resale price as the age increase.
- Model B has a greater initial resale price.
- Model B reduces in resale price by £95 per year.