

Statistics GCSE**Paper 1**

Edexcel Foundation - 2026

Foundation Tier

Variant 4

1ST0/1F

Instructions

- Write all answers in the spaces provided.
- Answer all questions.
- You must show all your working.
- There may not be enough space to show all your working out.

Information

- This is a practise paper to aid your revision for your exams.
- This site, and all that work on it, have no affiliation or relationship with any exam board.
- This site is not endorsed by any company or charity, unless we state otherwise.

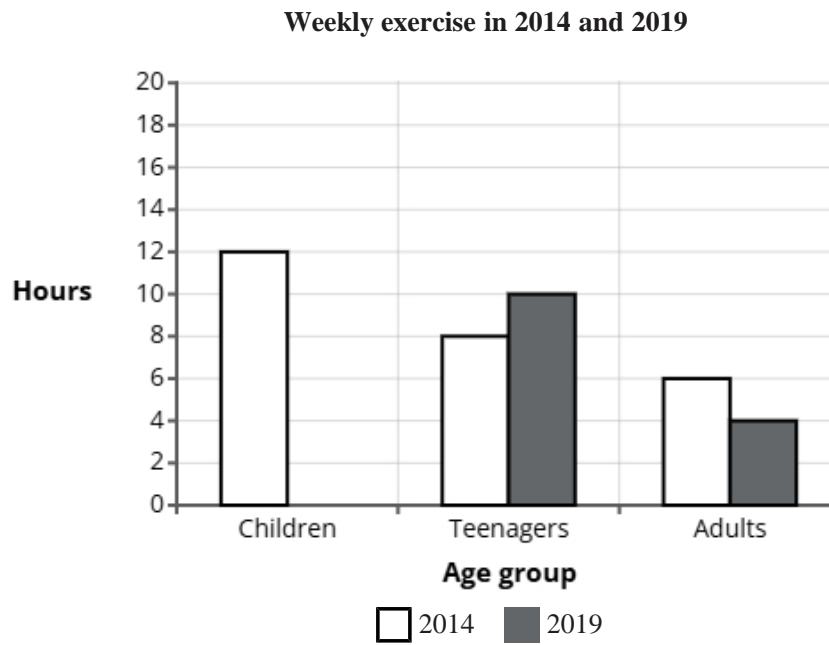
Copyright

- Sharing this PDF is strictly forbidden unless you have the author's permission.
- This paper is only authorised to be used by the person who has bought it.
- You may not store this online or any shared area such as an intranet.
- Please contact us if you have any queries.

Advice

- You can get support for all these questions at our website: www.statsgcse.com
- This paper and more are available on our site with questions that change subtly after each attempt.
- Good luck!

- 1 The comparative bar chart compares the average number of hours spent exercising per week by three age groups in 2014 and 2019.

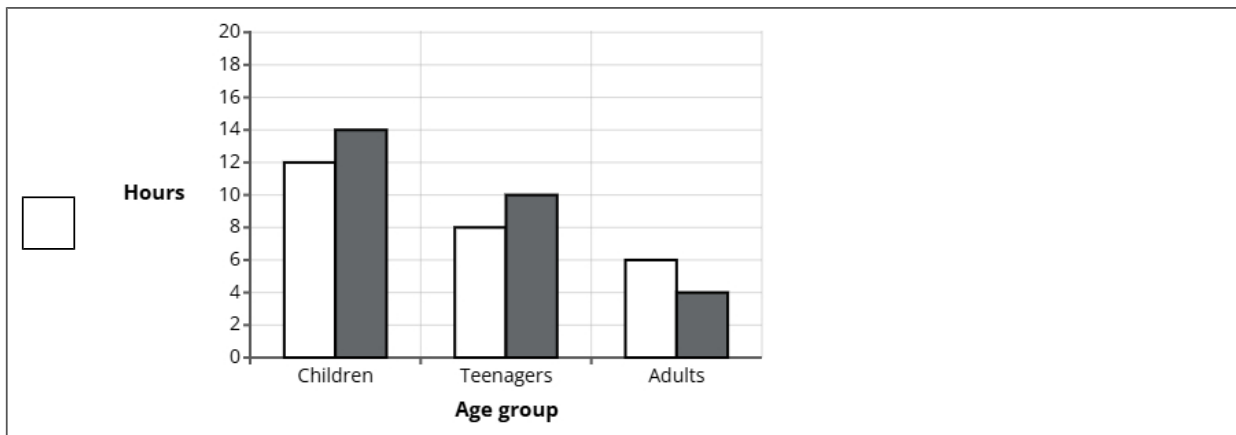
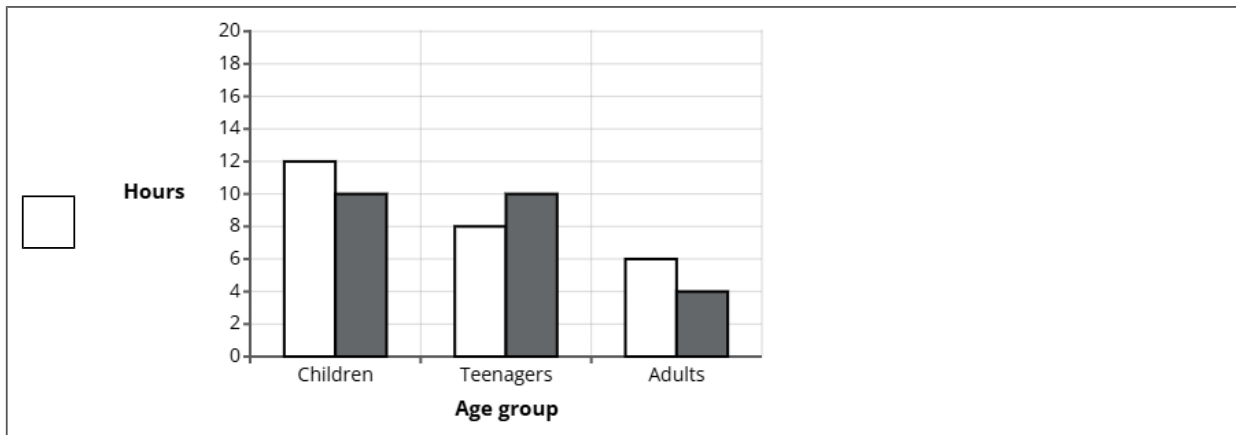
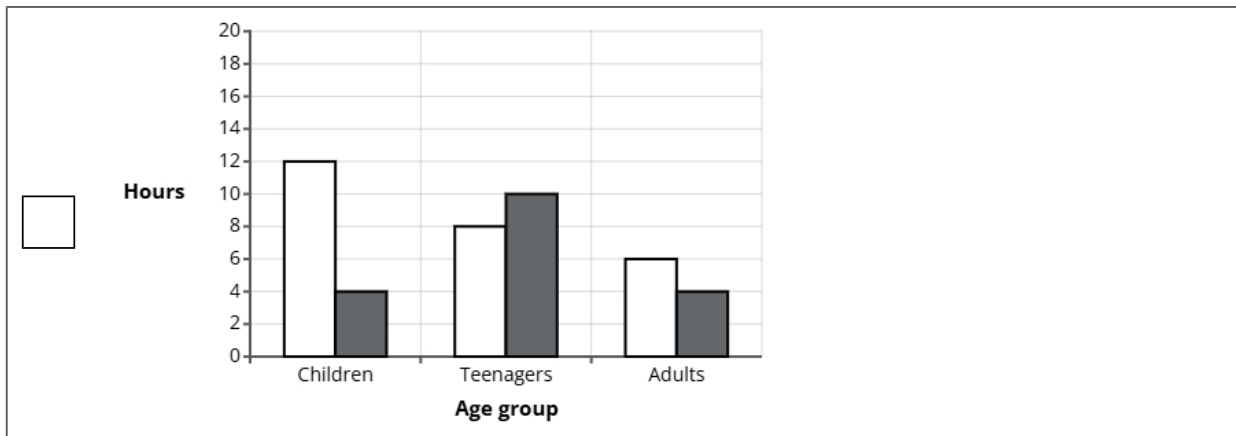
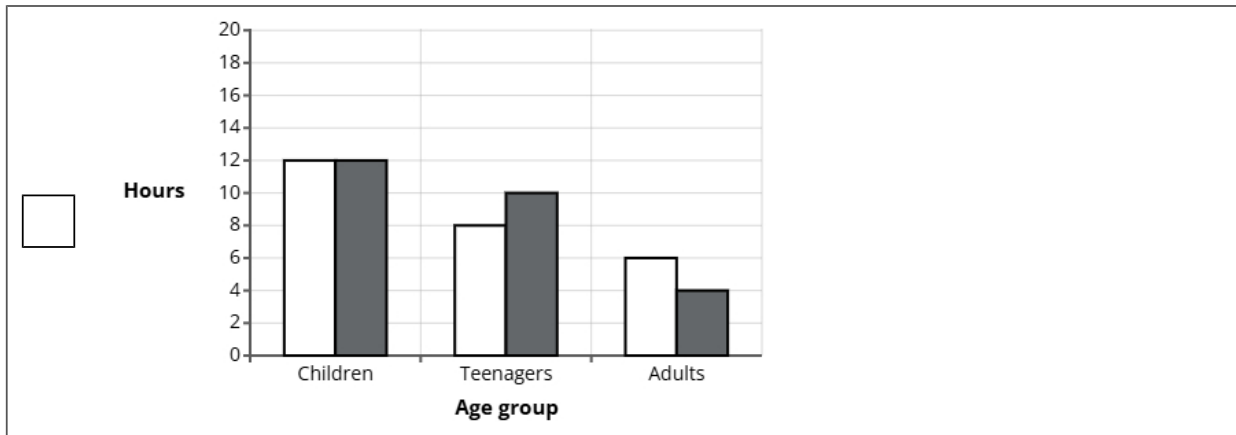


In 2019, children exercised for 14 hours.

- (a) Complete the comparative bar chart for children.

(1 mark)

Select the correct answer.



(b) Find how many more hours were spend exercising by teenagers than adults in 2014.

(2 marks)

2014 is shown with an unshaded bar

(c) Compare the the number of hours exercised by children, teenagers and adults in 2014.

(2 marks)

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

- Adults had the least hours of exercise.
- Teenagers had the least hours of exercise.
- Children had the most hours of exercise.
- Adults had the most hours of exercise.

(d) The data displayed in the comparative bar chart is an example of quantitative data.

Explain what is meant by quantitative data.

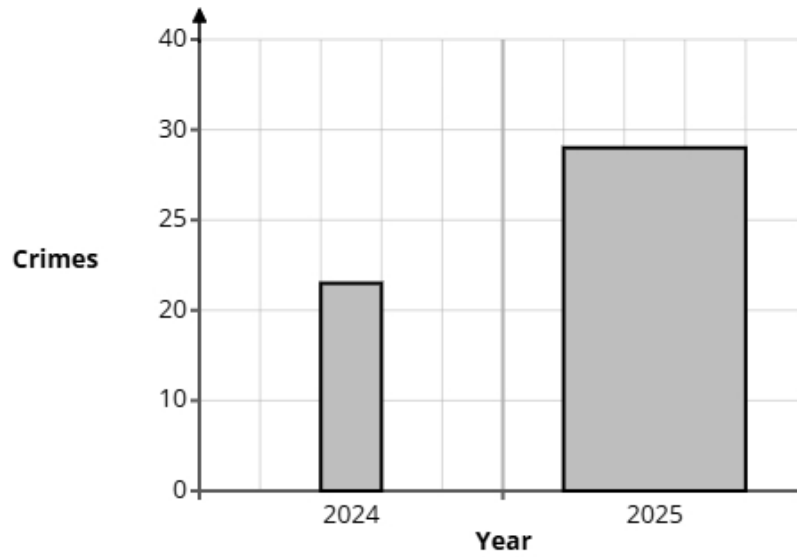
(1 mark)

Select **one** box.

- Quantitative data is data that focuses on detailed descriptions of events.
- Quantitative data is information that uses categories rather than numbers.
- Quantitative data is data that explores personal meanings and experiences.
- Quantitative data is numerical data that can be counted or measured.

2 A local newspaper editor sourced data to compare crime levels in a village in 2024 and 2025.

A bar chart is drawn from the information.



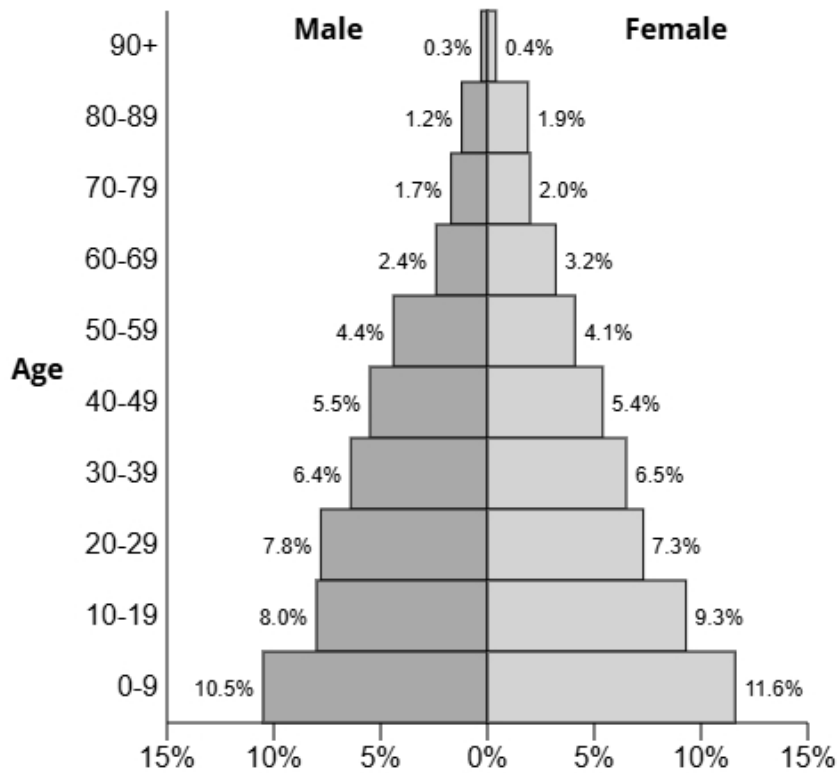
State **two** reasons why the bar chart could be misleading.

(2 marks)

Select **two** boxes.

- The vertical axis scale does not go up in equal steps.
- The bar width for 2025 is bigger than 2024.
- The bars have different areas.
- The heights are wrong.

3 The population pyramid below shows the percentage of males and females in each age group for the town Marshcombe.



(a) Write down the percentage of females in the age group 20-29.

(1 mark)

Find 20-29 on the population pyramid and read off the number on the right-hand side (females).

_____ %

(b) Find the age group for males that has 6.4% of the population.

(1 mark)

Select **one** box.

60-69

30-39

40-49

50-59

(c) Find the age group that has 17.3% of the population.

(1 mark)

Select **one** box.

40-49

20-29

30-39

10-19

(d) Compare the percentage of the population aged 50-79 between males and females.

(1 mark)

Select **one** box.

There are more females.

They are both the same.

There are more males.

(e) Give a reason why the sum of all the percentages is 99.9% and not 100%.

(1 mark)

Select **one** box.

Some of the population may not have been counted.

They are percentages not amounts.

The figures are wrong.

The figures have been rounded.

4 A hospital is planning to introduce a new appointment booking system.

Sophia wants to carry out a survey to find out what all patients think about the proposed change.

Sophia thinks that she should take a sample rather than a census.

(a) Sophia has decided to use the electoral register as a sampling frame.

State one problem Sophia may have using the electoral register as a sampling frame.

(1 mark)

Select **one** box.

The electoral register would also include people's addresses.

Bias.

Sophia may have difficulty gaining access to the electoral register.

There will be too many names.

(b) Sophia is writing a plan for the investigation into patients' opinions on the new appointment booking system.

Write down what Sophia should include in the plan and explain why each of the things is appropriate.

You should include:

- + a sampling method
- + a question Sophia should ask in the questionnaire
- + a statistical diagram to show the results of the survey.

(6 marks)

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- This will ensure that the sample is representative of the population.
- Sophia should use a case study.
- This will ensure that more students are asked.
- Sophia should use stratified sampling.

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- The question is clear and unbiased, avoiding leading students to a particular answer.
- A question could be:
Why do you think the appointment booking system needs to improve?
- The question is open so will be easier and quicker to analyse.
- A question could be:
 How satisfied are you with the current appointment booking system?
 Very Satisfied Satisfied Neutral Unsatisfied Very Unsatisfied

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- A tally chart can be used to display the data.
- A bar chart can be used to display the data.
- This is because it can be used also to collect the data.
- This is because it shows frequencies and allows for visual comparisons.

5 The council leader is reviewing the frequency of a local bus service.

The council wants to collect opinions from residents.

They plan to give a questionnaire to people using the bus service on one weekday and ask them to return it to the bus driver the following day.

(a) Describe the meaning of the term 'simple random sample'.

(1 mark)

Select **one** box.

- Choosing 10 employees from different departments to ensure variety.
- Taking the first 10 employees who volunteer to participate.
- Selecting people who have worked the longest.
- All members of the population have the same likelihood of selection.

(b) Assess the council leader's plan to get the opinions of the residents.

(3 marks)

Select **three** boxes.

- No residents will return the questionnaires.
- The residents are only being asked on one day.
- Questionnaires are a good way to collect opinions.
- The residents are not being selected at random.
- Not effective.

(c) Here is an open question that the council leader is considering for the questionnaire.

What do you think about the local bus service?

Give one reason why this is not a good question.

(1 mark)

Select **one** box.

- The question is not relevant.
- It is not a closed question.
- It may not be clear what type of answer is required.
- There may be issues with privacy.

(d) Design a suitable closed question for the council leader to use on her questionnaire so that she can decide how many busses to run per hour on this route.

(2 marks)

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

- Clean
- Dirty
- Busy
- Empty
- How would you describe the quality of this bus?
- How many busses per hour do you think should run on this route?
 - 1 to 2 buses per hour
 - 3 to 4 buses per hour
 - 5 to 6 buses per hour
 - More than 6 buses per hour

- (e) When the council leader has designed her questionnaire, she decides to pre-test it by using a pilot survey with a small sample of residents.

Select **two** reasons why she should conduct a pilot survey.

(2 marks)

Select **two** boxes.

- A pilot survey will include more people.
- A pilot survey will give more accurate data.
- A pilot survey will check questions are inoffensive.
- A pilot survey will test questions are working as intended.

- 6 A health researcher is investigating how much time adults spend walking each week.

She plans to ask a sample of 40 adults to record the number of hours they spend walking in one week.

Each person will write their total on a form provided.

Describe one problem the health researcher might face in the statistical enquiry process due to non-response or unexpected results, and explain how she could address this issue.

(2 marks)

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

- They may not give the answers in the same units, such as hours.
- The adult could lie.
- To address this she could tell them to give their answers to the nearest hour.
- To address this she could ask more young adults.

- 7 Jamie collected data on 11 flights, recording the distance (in hundreds of miles) and the price (in hundreds of pounds) of each flight. He represented his findings in the scatter diagram below.



- (a) One of the 11 flights has a price of £550.
For this flight, write down the distance.

(1 mark)

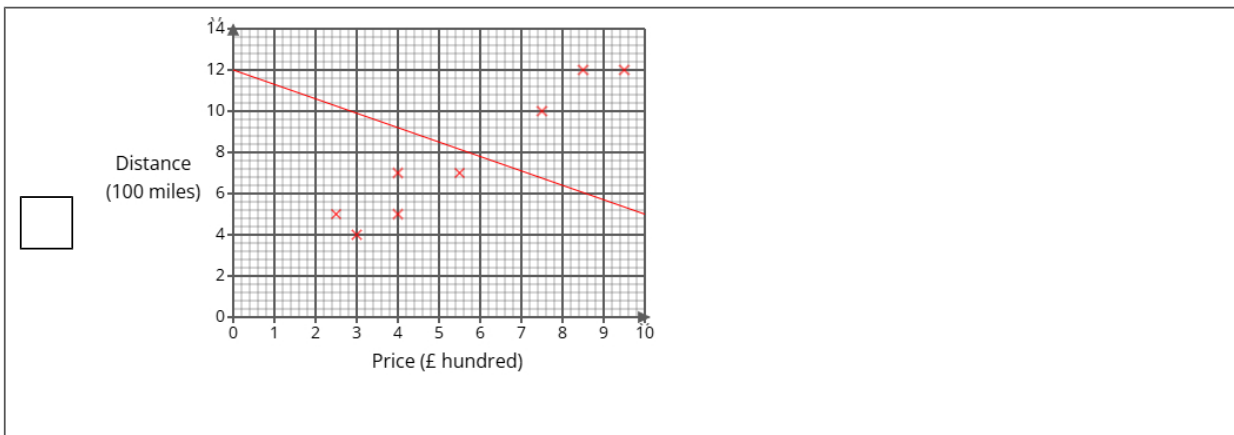
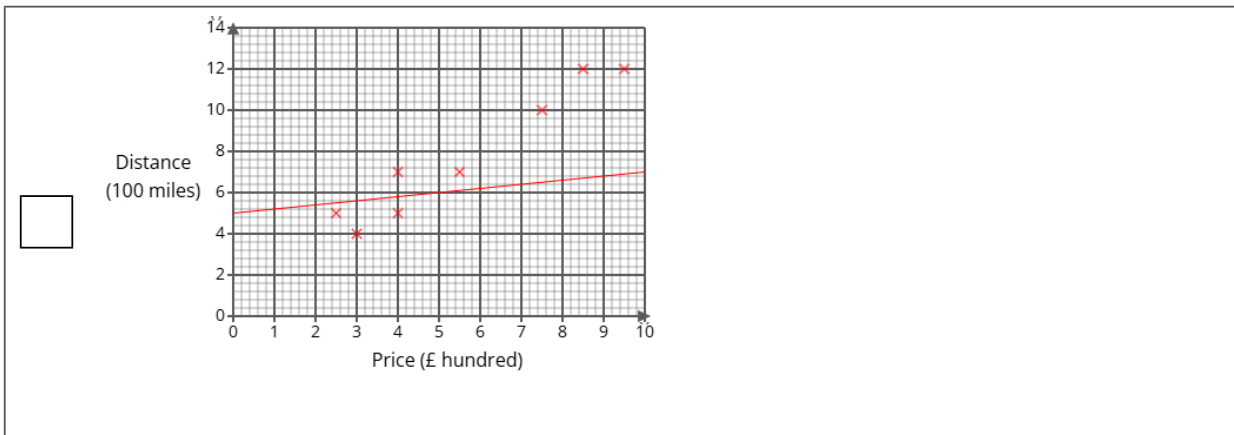
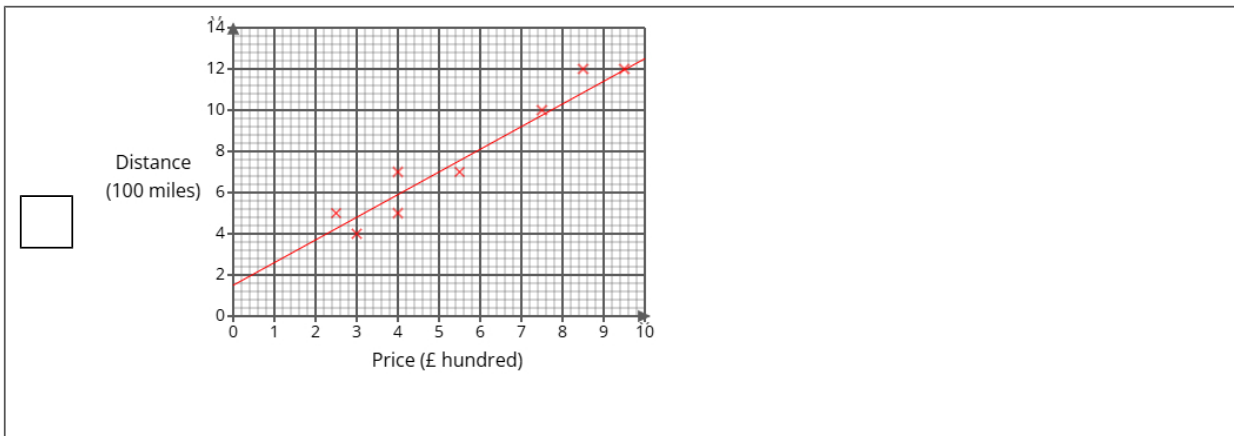
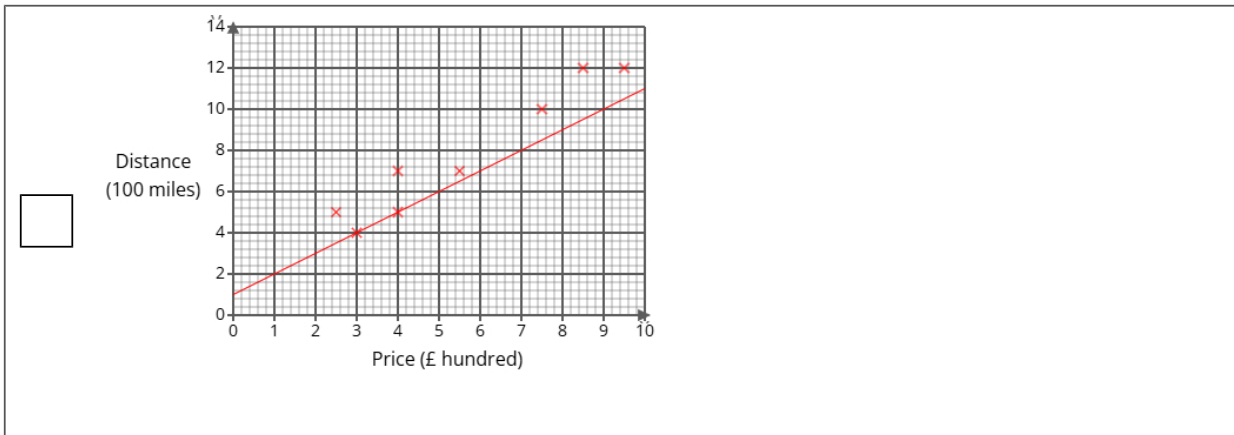
Find the cross on the scatter graph that is at [s036]550 on the x-axis (the bottom axis), then read off the value from the y-axis (the side axis).

_____ miles

(b) Draw a line of best fit on the scatter diagram.

(1 mark)

Select the correct answer.



(c) Describe and interpret the type of correlation shown by the scatter diagram.

(3 marks)

Number the **two** correct statements in the correct order (**three** statements are incorrect).

- The correlation is positive and
- There is no correlation but it is
- The correlation is negative and
- weak
- strong

Select **one** box.

- As the price increases the journey distance decreases.
- A journey that has a high price will have a low journey distance.
- A journey that has a high price will have a high journey distance.
- As the price increases the journey distance increases.

(d) An airline has announced a new route at a price of £2000.

Jamie is planning on using the line of best fit on the scatter diagram to predict the distance of the flight.

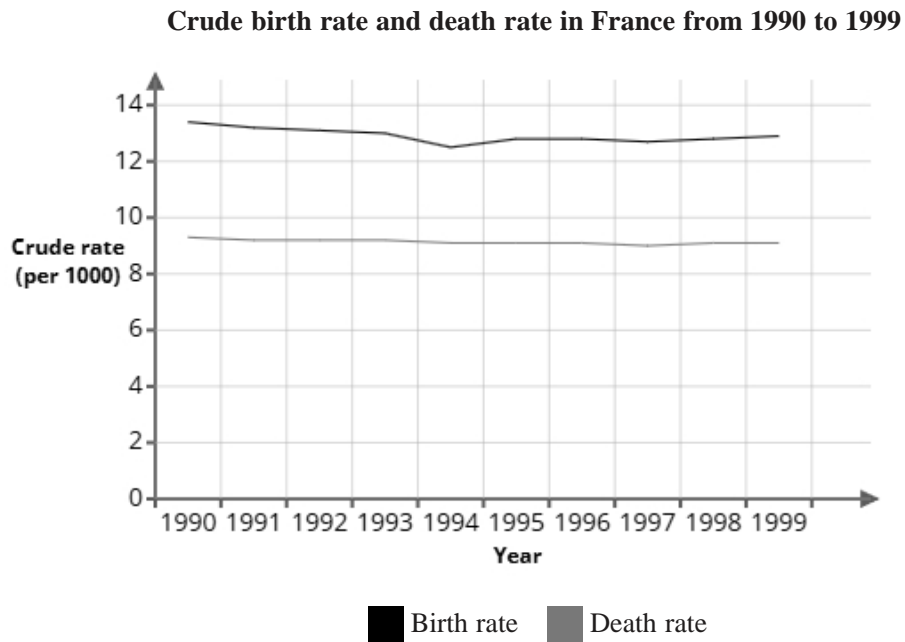
Explain whether or not it is appropriate to use the line of best fit for this prediction.

(2 marks)

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- This is appropriate
- because the point is after the data and the trend may not continue.
- This is not appropriate
- because the trend will continue.

8 The graph shows the crude birth rate and death rate in France from 1990 to 1999



Liam uses the information in the graph to conclude:

"The total population in France has increased from 1990 to 1999"

(a) Explain how the information in this graph supports Liam's conclusion.

(1 mark)

Select **one** box.

- The death rate is higher than the birth rate.
- The graph shows that France is a large country.
- In 1990, the birth rate is roughly 4 bigger than the death rate.
- The birth rate is higher than the death rate.

(b) Give a reason why Liam's conclusion might **not** be correct.

(1 mark)

Select **one** box.

- Liam's conclusion does not take into account people moving out of the country.
- Liam's conclusion does not take into account where in France these births happened.
- The data in the graph may be inaccurate.
- Liam's conclusion does not take into account the average age of the population.

(c) In 2000, the population of France was 59 934 350.

There were 638 345 recorded births.

Using the formula below, calculate the crude birth rate in 2000.

Give your answer correct to 1 decimal place.

$$\text{crude birth rate} = \frac{\text{number of births} \times 1000}{\text{total population}}$$

(2 marks)

Substitute each of the values into the crude birth rate formula.

9 Noah investigates the reaction times (in milliseconds) of 120 people taking a driving test.

The times range from 203 ms to 281 ms.

Noah considers using one of the two possible grouped frequency tables for the results, Table A or Table B, shown below.

Table A

Time (t ms)	Frequency
$170 < t \leq 200$	0
$200 < t \leq 230$	31
$230 < t \leq 260$	67
$260 < t \leq 290$	22
$290 < t \leq 310$	0

Table B

Time (t ms)	Frequency
$200 < t \leq 220$	14
$220 < t \leq 240$	40
$240 < t \leq 260$	44
$260 < t \leq 280$	20
$280 < t \leq 300$	2

(a) Give **two** advantages of using grouped data rather than raw data.

(2 marks)

Select **two** boxes.

- Grouped data doesn't require any calculation, as it displays frequencies directly.
- Grouped data provides a more accurate reflection of trends.
- Grouped data can help to spot patterns in the data.
- Grouped data is easier to represent on graphs.
- Grouped data helps avoid using averages.

(b) Give **one** disadvantage of using grouped data rather than raw data.

(1 mark)

Select **one** box.

- Grouped data is much more difficult to read.
- Grouped data cannot be drawn on a graph.
- Grouped data can only calculate estimates of statistical values.
- Grouped data cannot be compared.

(c) Noah feels that Table B gives more detail than Table A about the results.

Assess the appropriateness of Noah's claim.

(2 marks)

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

- In Table A, the data goes from 170 to 210, showing a much wider range of data.
- In Table B, some data could have been less than 200 or more than 300, but would not be shown.
- Noah's claim is justified.
- In Table B, the table starts at 200 and the lowest value is 203 ms and ends at 300 with the highest value at 281 ms.
- Noah's claim is not justified.
- In Table A, all the data is concentrated into three groups.

(d) Noah wants to work out the average reaction times of the 120 people taking a driving test.

He decides to use Table B.

Calculate the average reaction times of the 120 people taking a driving test, giving your answer to 1 decimal place.

(3 marks)

Add midpoint and ft columns onto the table.

Then find the sums of the f and ft .

Find the mean by $\frac{\sum ft}{\sum f}$

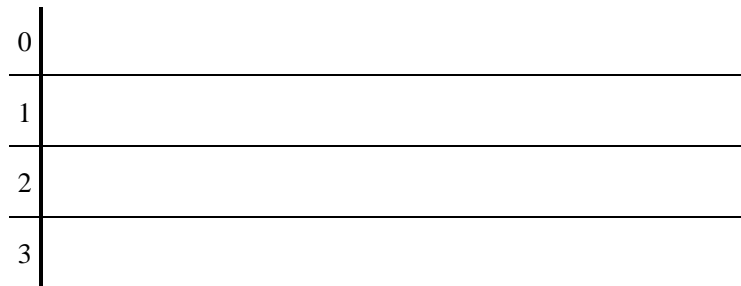
_____ ms

10 23 female adults were asked to skip with a rope continuously for 1 minute and the number of jumps was measured.

Here are the results.

25	21	11	34	14	10	27	4
0	23	25	30	14	27	4	14
24	7	6	32	37	8	7	

(a) Complete the stem and leaf diagram for the data.



Key: 1|0 = 10

(2 marks)

Select the correct answer.

	0	0 4 4 6 7 7 8
	1	1 3 4 5 5 7 7
<input type="checkbox"/>	2	0 1 4 4 4
	3	0 2 4 7

	0	0 4 4 6 7 7 8
	1	0 1 4 4 4
<input type="checkbox"/>	2	0 2 4 7
	3	1 3 4 5 5 7 7

	0	0 4 4 6 7 7 8
	1	0 1 4 4 4
<input type="checkbox"/>	2	1 3 4 5 5 7 7
	3	0 2 4 7

	0	0 1 4 4 4
	1	0 4 4 6 7 7 8
<input type="checkbox"/>	2	1 3 4 5 5 7 7
	3	0 2 4 7

(b) Work out the interquartile range from the data.

(2 marks)

Find the lower quartile using $\frac{1}{4}(n + 1)^{\text{th}}$

Find the upper quartile using $\frac{3}{4}(n + 1)^{\text{th}}$

Find the interquartile range

IQR = upper quartile – lower quartile

(c) A group of 23 male adults were also measured.

The results from the male adults had a median of 18 and an interquartile range of 17.

Sophie thinks that these results show that males are better at skipping than females.

State whether you agree with Sophie and give reasons why.

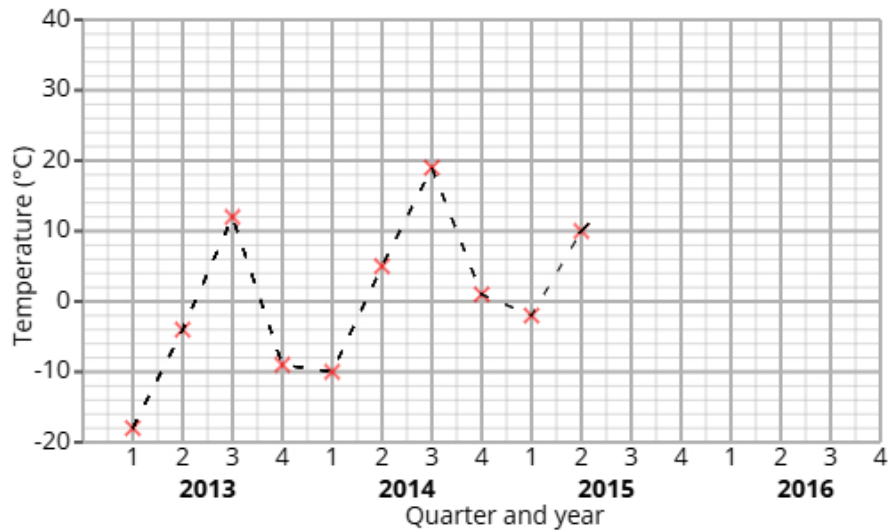
(3 marks)

Median for female adults = _____

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

- The median for the females is higher than the median of the males.
- No, Sophie's conclusion is supported by the data.
- Yes, Sophie's conclusion is supported by the data.
- The median for the males is higher than the median of the females.

11 The time series graph shows information about the the temperature of Eastern Galm from 2013 to 2015.



Chloe calculates the 4-point moving averages from the time series graph, which are shown below.

-5 -3 -1 1 4 6 7

(a) Identify and interpret in context one example of seasonality displayed in the time series graph.

(2 marks)

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- which shows that the lowest temperature is in the summer.
- which shows that the highest temperature is in the summer.
- The lowest values are in Q2
- The greatest values are in Q3

(b) Explain why a 4-point moving average is appropriate.

(1 mark)

Select **one** box.

It shows skew within the data.

It shows correlation.

A 4-point moving average gives us more data.

A 4-point moving average captures cyclic patterns that repeat every four quarters.

12 A fair 5-sided spinner is numbered 1, 2, 3, 4, 5.

A fair 3-sided spinner is numbered 1, 2, 3.

The spinners are used to play a game. Both spinners are spun and the total score is recorded.

		3-sided spinner		
		1	2	3
5-sided spinner	1	2	3	
	2	3		
	3			
	4			
	5			

The game is won when the total is at least 6.

Sofia plays the game once.

(a) Complete the sample space diagram.

(2 marks)

Select the correct answer.

		3-sided spinner		
		1	2	3
5-sided spinner	1	2	3	3
	2	3	3	4
	3	3	4	5
	4	4	5	6
	5	5	6	7

		3-sided spinner		
		1	2	3
5-sided spinner	1	2	3	3
	2	3	4	6
	3	3	6	9
	4	4	8	12
	5	5	10	15

		3-sided spinner		
		1	2	3
5-sided spinner	1	2	3	4
	2	3	4	5
	3	4	5	6
	4	5	6	7
	5	6	7	8

(b) Find the probability that Sofia wins the game.

(2 marks)

Find all the numbers in the table that are 6 or larger

Put this number as the numerator and total amount of numbers as the denominator

$$\text{probability} = \frac{\text{6 or larger}}{\text{total outcomes}}$$

- 13** Noah organises two coding bootcamps, Bootcamp 1 and Bootcamp 2, to teach basic Python programming. He wants to compare the two bootcamps to see which teaches Python better. The table shows number of participants who passed and failed the coding assessment.

	Passed	Failed	Total
Bootcamp 1	24	28	52
Bootcamp 2	15	50	65

(i) Find the relative risk of failing the coding assessment having been in Bootcamp 1 compared to Bootcamp 2.

(ii) Give an interpretation of your answer to part (i).

(4 marks)

Write your answer as a decimal.

Select **one** box.

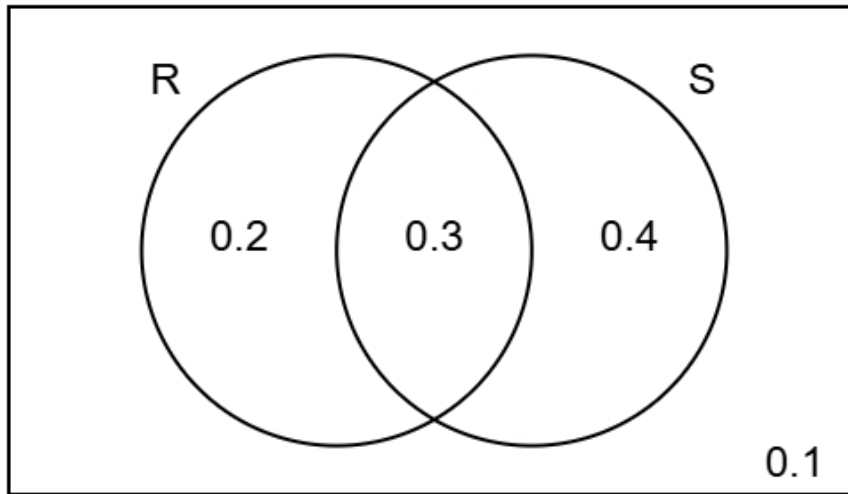
More people failed the coding assessment in Bootcamp 1 than in Bootcamp 2.

The risk of failing the coding assessment having taken Bootcamp 1 is lower than the risk of failing the coding assessment having taken Bootcamp 2.

Less people failed the coding assessment in Bootcamp 1 than in Bootcamp 2.

The risk of failing the coding assessment having taken Bootcamp 1 is greater than the risk of failing the coding assessment having taken Bootcamp 2.

- 14 The Venn diagram shows information about the probabilities of two events occurring.
The events are labelled as R and S.



- (a) Find the probability of event S happening.

(1 mark)

Add the probabilities in the circle marked S together

Leave your answer as a decimal.

- (b) Find $P(R \text{ and } S)$

(1 mark)

$P(R \text{ and } S)$ is shown in the overlap of the Venn diagram

Leave your answer as a decimal.

(c) Find $P(S | R)$

(2 marks)

Use the formula to find $P(S | R)$

$$P(S | R) = \frac{P(R \text{ and } S)}{P(R)}$$

(d) Two different events events D and E are independent.

$$P(D) = 0.2$$

$$P(E) = 0.6$$

Find $P(D \text{ and } E)$

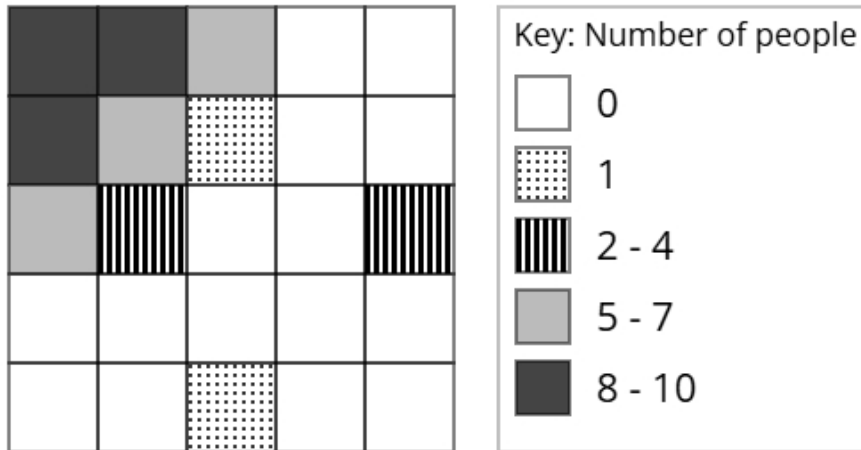
(2 marks)

Use the formula for independent events

$$P(D \text{ and } E) = P(D) \times P(E)$$

Leave your answer as a decimal.

- 15** The choropleth map below represents an airport terminal that has been divided into 25 squares of equal area. Daniel has collected data about the popularity of different parts of the airport terminal. The number of people recorded in each square on one Monday morning is shown.



- (a) Calculate an estimate of the total number of people that were recorded on Monday.

(3 marks)

Find the midpoints for the groups.

Multiply each key with amount of squares and add them up.

(b) Daniel would like to open a bubble tea booth in the airport terminal.

After analysing the data, he decides that he should open the bubble tea booth in the corner of the airport terminal shown at the top left of the choropleth map.

Using the information in the choropleth map, assess the validity of Daniel's conclusion.

(2 marks)

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

- because there were less people at the top left of the airport terminal.
- Daniel's comment is valid
- Daniel's comment is not valid
- because there were more people at the top left of the airport terminal.

(c) Fatima argues that the method used by Daniel to collect the data is not appropriate for reaching a reliable conclusion.

Assess whether Fatima's argument is correct and give a reason.

(1 mark)

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

- Fatima is not correct
- because the data was only collected on one Monday.
- Fatima is correct
- because there was a large amount of data collected.