

**Statistics GCSE****Paper 1**

Edexcel Foundation - 2026

Foundation Tier

Variant 1 (same as video)

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.

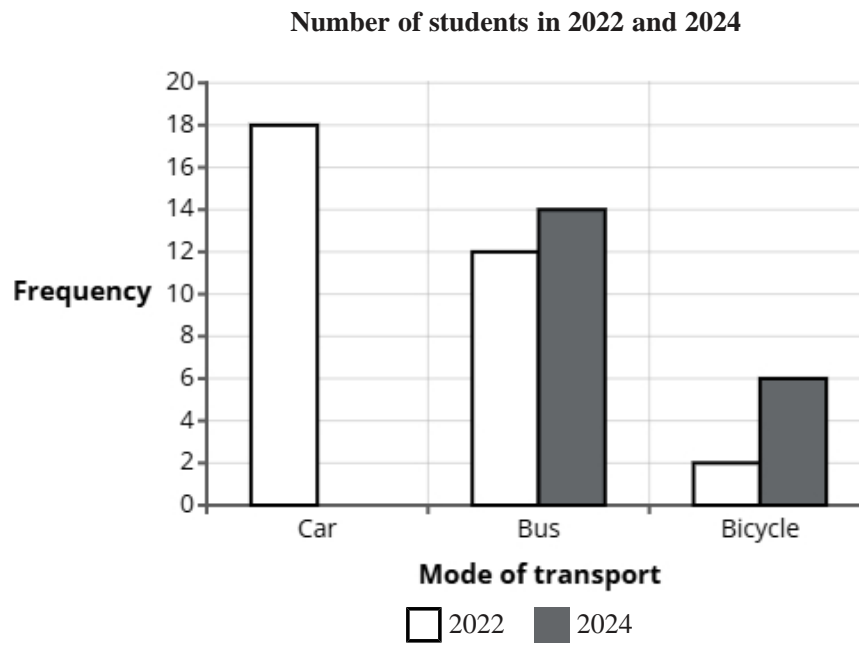
**Copyright**

- 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](http://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 number of students who travel to a school by car, bus and bicycle in 2022 and 2024.

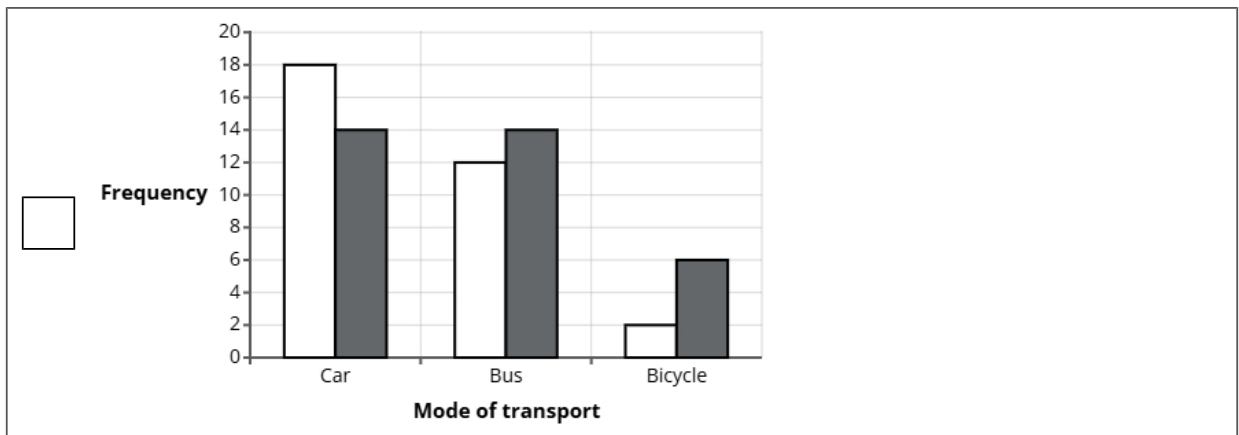
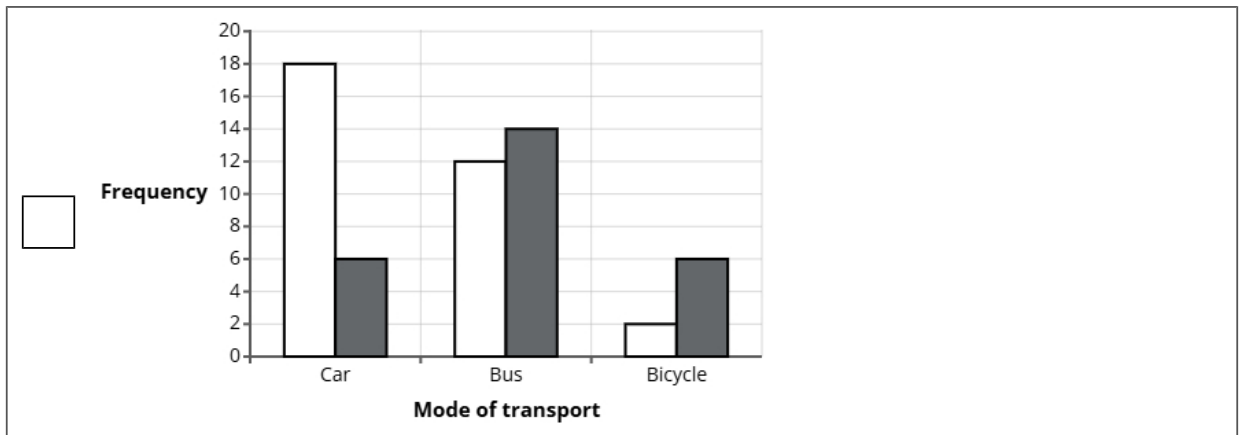
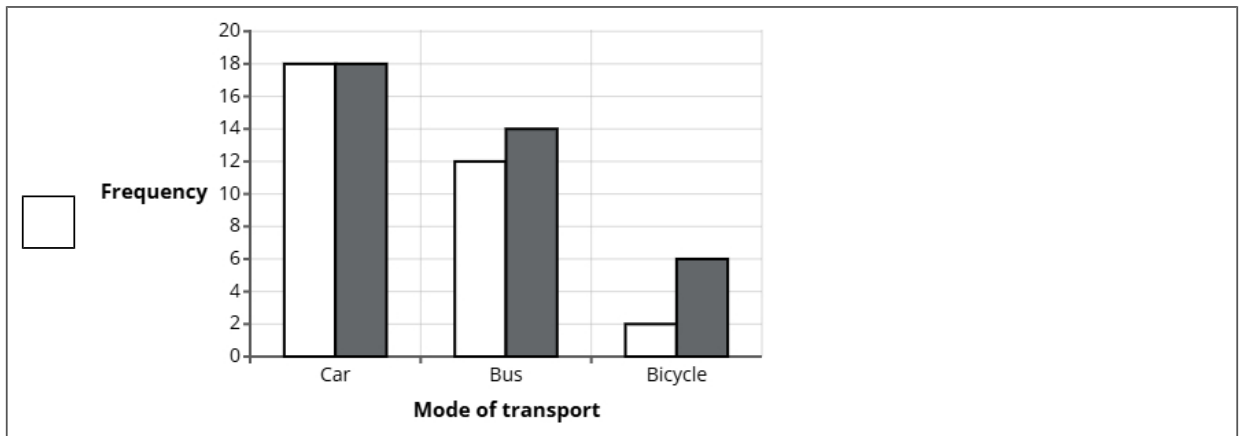
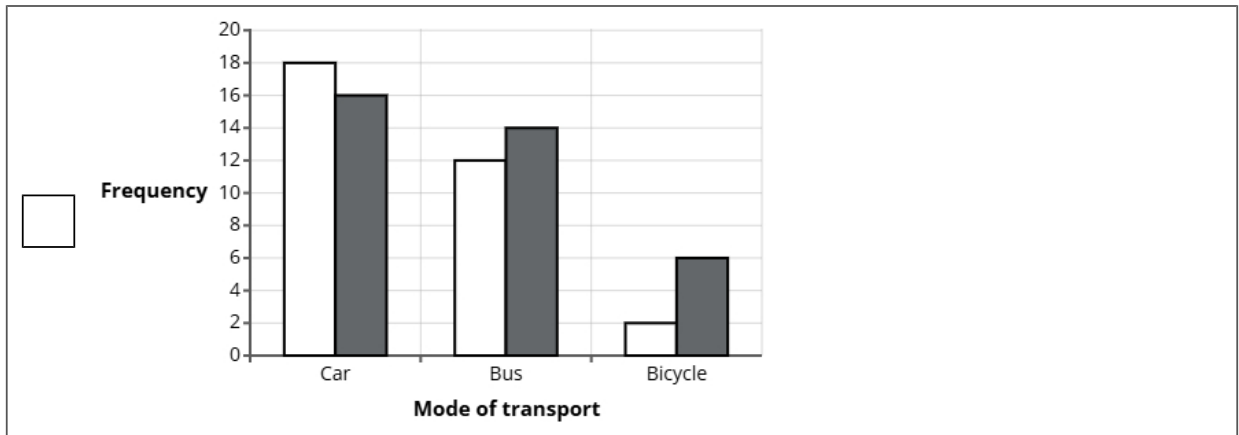


In 2024, there were 16 students who travelled to school by car.

- (a) Complete the comparative bar chart for students who travelled by car.

(1 mark)

Select the correct answer.



(b) Find how many more students travelled by bus than bicycle in 2022.

(2 marks)

2022 is shown with an unshaded bar

(c) Compare the number of students who took the car, bus and bicycle in 2022.

(2 marks)

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

- The bicycle was the least popular method of transport.
- The car was the most popular method of transport.
- The bus was the most popular method of transport.
- The car was the least popular method of transport.

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

Explain what is meant by quantitative data.

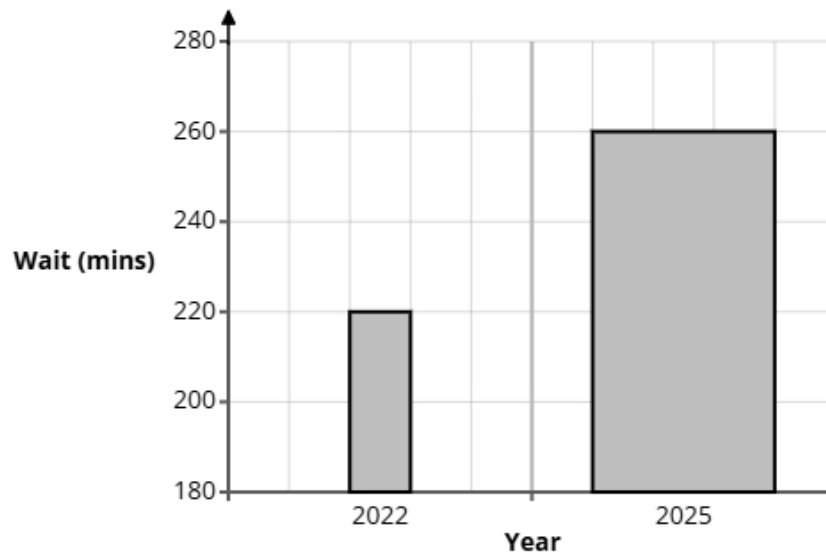
(1 mark)

Select **one** box.

- Data that describes qualities or characteristics using words.
- Data that can be measured or counted and is expressed using numbers.
- Data that explains meanings or reasons behind behaviours.
- Data collected through opinions and personal experiences.

- 2 A journalist carried out research into hospital waiting times at a hospital in 2022 and 2025 to highlight concerns about local healthcare.

A bar chart is drawn from the information.



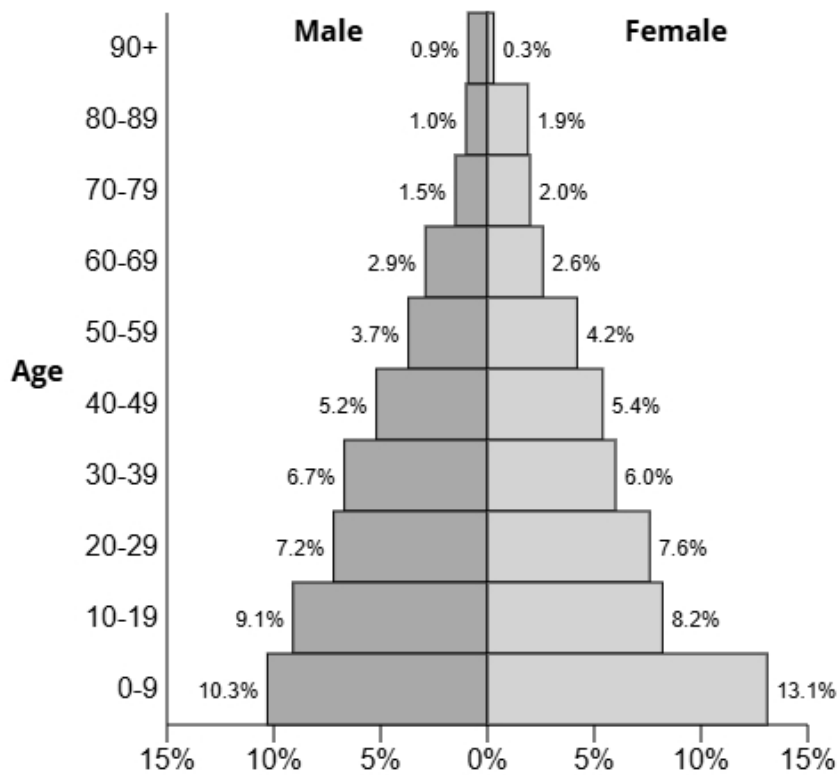
State **two** reasons why the bar chart could be misleading.

(2 marks)

Select **two** boxes.

- There is no title for the graph.
- The vertical axis on the graph does not begin at 0.
- The bar width for 2025 is bigger than 2022.
- The bars have different areas.

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



(a) Write down the percentage of females in the age group 50-59.

(1 mark)

Find 50-59 on the population pyramid and read off the number on the right-hand side (females).

\_\_\_\_\_ %

(b) Find the age group for males that has 6.7% of the population.

(1 mark)

Select **one** box.

50-59

40-49

30-39

20-29

(c) Find the age group that has 14.8% of the population.

(1 mark)

Select **one** box.

20-29

40-49

30-39

50-59

(d) Compare the percentage of the population aged 30-59 between males and females.

(1 mark)

Select **one** box.

They are both the same.

There are more females.

There are more males.

(e) Give a reason why the sum of all the percentages is 99.8% and not 100%.

(1 mark)

Select **one** box.

- It is incomplete.
- They have been rounded.
- There are gaps in the ages.
- The figures are wrong.

4 A school is considering introducing a new lunch menu.

Daniel is going to conduct a survey to gather opinions from everyone in the school.

Daniel thinks that he should take a sample rather than a census.

(a) Daniel has decided to use the class registers as a sampling frame.

State one problem Daniel may have using the class registers as a sampling frame.

(1 mark)

Select **one** box.

Unreliable.

The class registers would also include dates of birth.

The staff will not be included.

There will be too many names.

(b) Daniel is writing a plan for the investigation into students' opinions on the new lunch menu.

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

You should include:

- + a sampling method
- + a question Daniel 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).

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

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

A question could be:

- How satisfied are you with the current school lunch options?  
 Very Satisfied  Satisfied  Neutral  Unsatisfied  Very Unsatisfied

A question could be:

- Why do you think the school lunch options need to improve?
- The question is open so will be easier and quicker to analyse.
- The question is clear and unbiased, avoiding leading students to a particular answer.

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

- This is because it shows frequencies and allows for visual comparisons.
- 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.

5 The head teacher of a secondary school is reviewing the amount of homework given each week.

The head teacher wants to collect the views of parents and guardians.

She plans to give a questionnaire to parents who attend the next parents' evening and ask them to return it to the school the next day.

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

(1 mark)

Select **one** box.

A sampling technique that ensures only the most relevant individuals are chosen.

A method that divides the population into groups before selecting individuals.

A sampling method where every individual in the population has an equal chance of being selected.

A method where individuals are selected based on convenience and availability.

(b) Assess the head teacher's plan to get the opinions of the parents.

(3 marks)

Select **three** boxes.

No parents will return the questionnaires.

Questionnaires are not good.

The parents are only being asked on one day.

The parents are not being selected at random.

Many parents will not return the questionnaire.

(c) Here is an open question that the head teacher is considering for the questionnaire.

What do you think about the amount of homework we give each week?

Give one reason why this is not a good question.

(1 mark)

Select **one** box.

- It is an open question.
- The question is not relevant.
- There are no response boxes.
- There may be issues with privacy.

(d) Design a suitable closed question for the head teacher to use on her questionnaire so that she can decide how much homework students are given each week.

(2 marks)

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

- 1 to 2 hours
- 3 to 4 hours
- 5 to 6 hours
- Less than 3 hours
- 3 to 5 hours
- 6 to 8 hours
- More than 8 hours
- Do you think the level of homework we give students is suitable?
- How much homework do you think students should be given each week?

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

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 give an idea of what the results may be.
- A pilot survey will help identify problems.

- 6 A youth worker wants to find out how much time teenagers spend exercising each week.

She plans to ask a sample of 50 young people at her youth club to record the number of hours they exercise in one week.

Each participant will write down their total exercise time for that week.

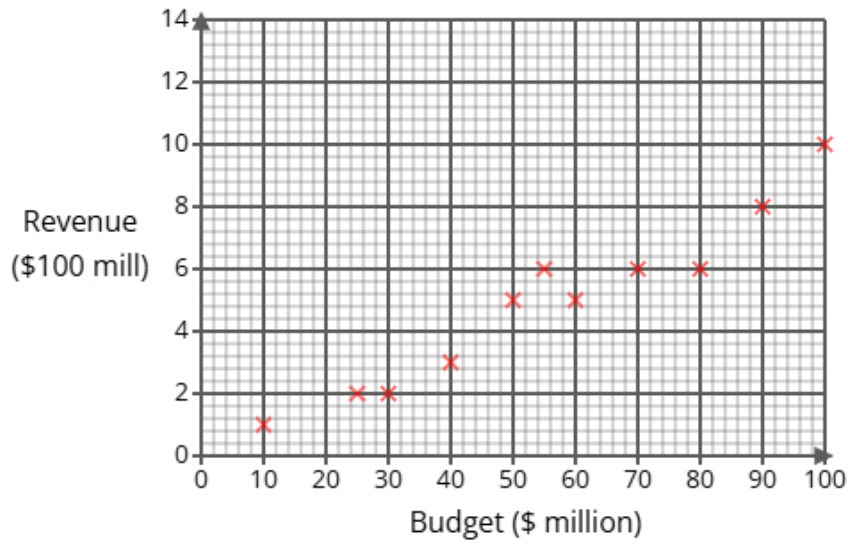
Describe one problem the youth worker might face in the statistical enquiry process due to non-response or unexpected results, and explain how she could address this issue.

(2 marks)

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

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

- 7 Liam gathered data on 11 movies, recording their production budget (in millions of dollars) and total box office revenue (in hundreds of millions). He represented his findings in the scatter diagram below.



- (a) One of the 11 movies has a budget of 55 million.

For this movie, write down its revenue.

(1 mark)

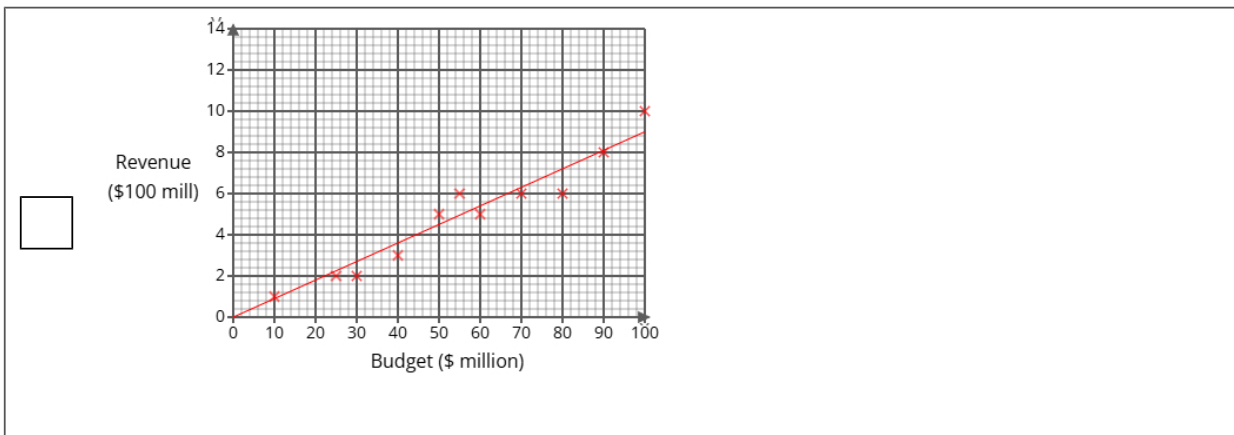
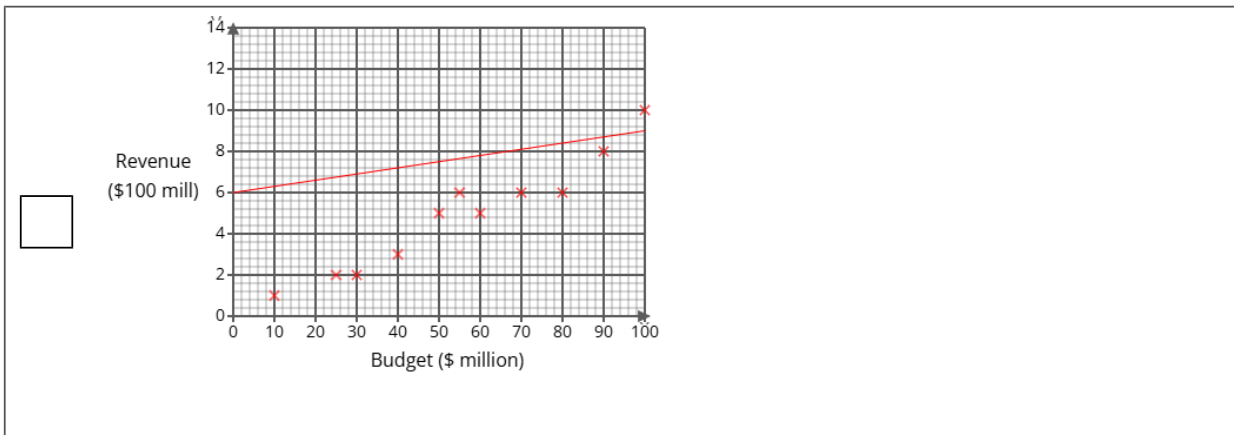
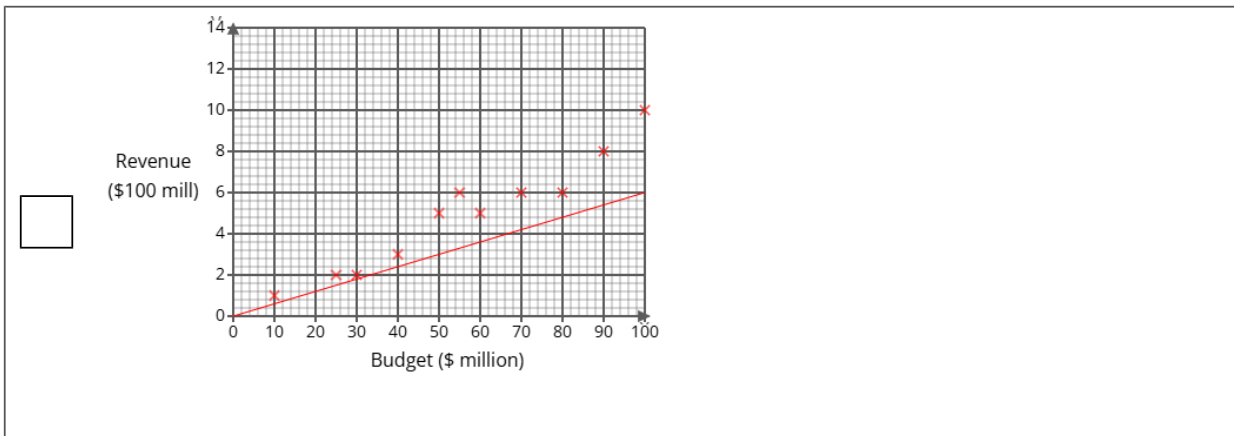
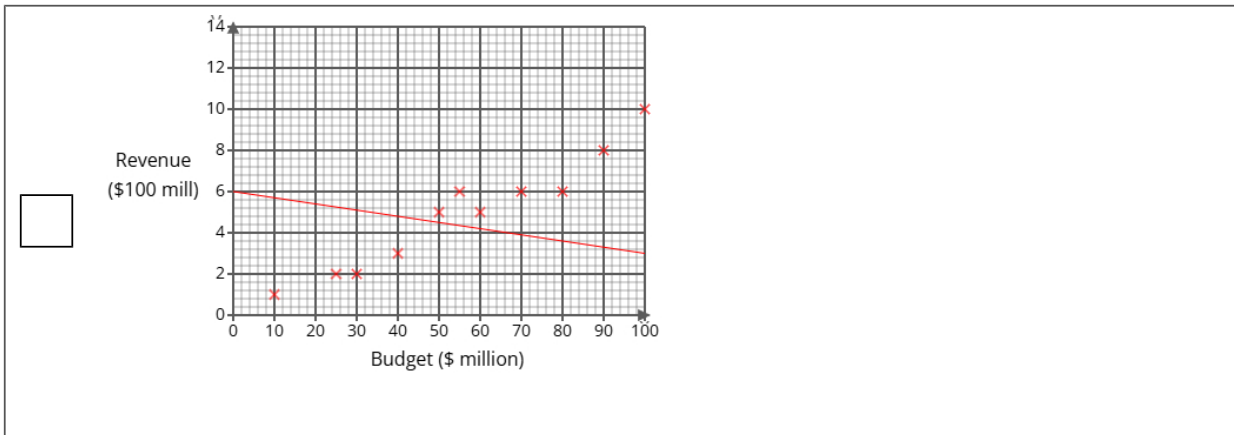
Find the cross on the scatter graph that is at \$55 million on the x-axis (the bottom axis), then read off the value from the y-axis (the side axis).

\$ \_\_\_\_\_

**(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).

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

Select **one** box.

- As the budget increases the revenue increases.
- A big budget movie will have a low revenue.
- A big budget movie will have a high revenue.
- As the budget increases the revenue decreases.

(d) A new movie will be releasing soon with a budget of \$250 million.

Liam is planning on using the line of best fit on the scatter diagram to predict the revenue of the movie.

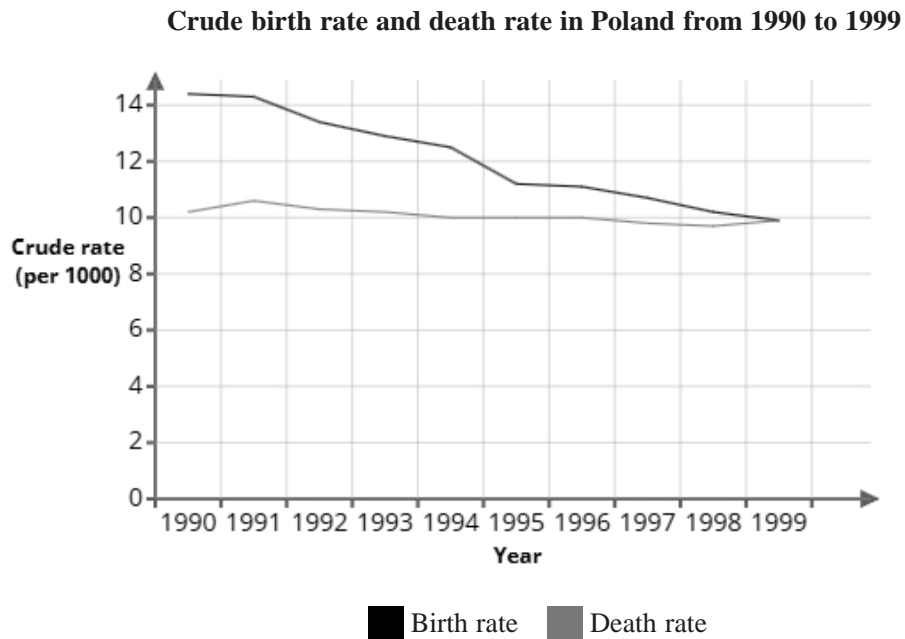
Explain whether or not it is appropriate to use the line of best fit for this prediction.

(2 marks)

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

- because this would be interpolation.
- This is not appropriate
- because this would be extrapolation.
- This is appropriate

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



Elena uses the information in the graph to conclude:

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

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

(1 mark)

Select **one** box.

- There are more births than deaths.
- The graph shows that Poland is a large country.
- In 1990, the birth rate is roughly 4 bigger than the death rate.
- There are more deaths than births.

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

(1 mark)

Select **one** box.

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

(c) In 2000, the population of Poland was 38 651 245.

There were 378 343 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 Sophia investigates the heights of 210 students in a school.

The heights range from 129 cm to 203 cm.

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

**Table A**

Height (h cm)	Frequency
$90 < h \leq 120$	0
$120 < h \leq 150$	37
$150 < h \leq 180$	147
$180 < h \leq 210$	26
$210 < h \leq 240$	0

**Table B**

Height (h cm)	Frequency
$120 < h \leq 140$	13
$140 < h \leq 160$	62
$160 < h \leq 180$	109
$180 < h \leq 200$	24
$200 < h \leq 220$	2

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

(2 marks)

Select **two** boxes.

- Grouped data makes it easier to process large amounts of data.
- Grouped data is quicker.
- Grouped data makes it easier to identify outliers.
- Grouped data can make it easier to compare different groups.
- Grouped data keeps the precision in the data.

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

(1 mark)

Select **one** box.

- Grouped data cannot be used for statistical tests.
- Grouped data will take longer to process large amounts of data.
- Grouped data will lose the detail in the data.
- Grouped data may introduce too much complexity when analysing trends.

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

Assess the appropriateness of Sophia's claim.

(2 marks)

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

- Sophia's claim is justified.
- In Table A, there are no students in two of the groups, so there is a much better spread of data.
- in Table B, each group has a smaller class width, showing more detail.
- in Table B, each group has a smaller class width, so detail is lost.
- In Table A, there are no students in two of the groups, so these groups are not needed.
- Sophia's claim is not justified.

(d) Sophia wants to work out the average height of the 200 students in the school.

She decides to use Table B.

Calculate the average height of the 200 students in the school, giving your answer to 1 decimal place.

(3 marks)

Add midpoint and  $fh$  columns onto the table.

Then find the sums of the  $f$  and  $fh$ .

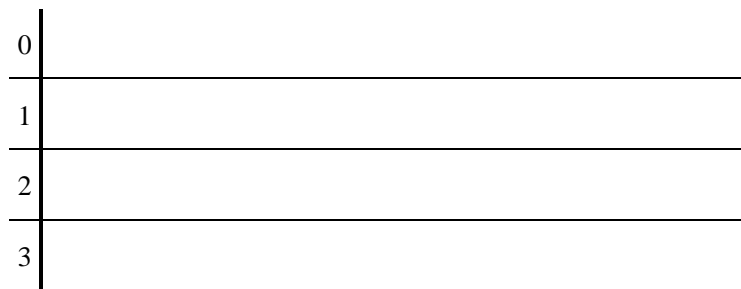
Find the mean by  $\frac{\sum fh}{\sum f}$

\_\_\_\_\_ cm

**10** 23 male volunteers were asked to hit targets with 50 darts and the number of targets hit was measured.  
Here are the results.

34	25	14	6	37	4	35	9
24	3	37	1	13	8	4	31
22	24	17	9	38	29	21	

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



**Key:** 1|3 = 13

(2 marks)

Select the correct answer.

	0	1 3 4 4 6 8 9 9
	1	1 2 4 4 5 9
<input type="checkbox"/>	2	3 4 7
	3	1 4 5 7 7 8

	0	1 3 4 4 6 8 9 9
	1	3 4 7
<input type="checkbox"/>	2	1 4 5 7 7 8
	3	1 2 4 4 5 9

	0	1 3 4 4 6 8 9 9
	1	3 4 7
<input type="checkbox"/>	2	1 2 4 4 5 9
	3	1 4 5 7 7 8

	0	3 4 7
	1	1 3 4 4 6 8 9 9
<input type="checkbox"/>	2	1 2 4 4 5 9
	3	1 4 5 7 7 8

(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 female volunteers were also measured.

The results from the female volunteers had a median of 23 and an interquartile range of 25.

Jimmy thinks that these results show that females are better at throwing darts than males.

State whether you agree with Jimmy and give reasons why.

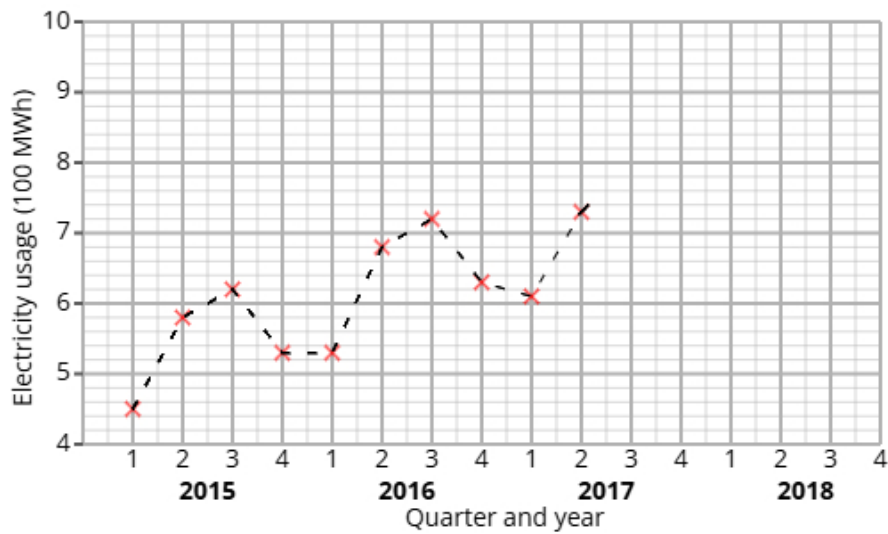
(3 marks)

Median for male volunteers = \_\_\_\_\_

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

- The median for the females is higher than the median of the males.
- No, Jimmy's conclusion is supported by the data.
- Yes, Jimmy'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 electricity usage at a school from 2015 to 2017.



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

5.5      5.7      5.9      6.2      6.4      6.6      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 less electricity is used by the school in the summer.
- The greatest values are in Q3
- which shows that more electricity is used by the school in the summer.
- The lowest values are in Q2

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

(1 mark)

Select **one** box.

- The pattern in the data repeats every four quarters.
- It allows us to predict future values.
- It eliminates all seasonality.
- 4-points is more detailed than annual data.

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

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

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

The game is won when the total is at least 5.

Aiden plays the game once.

- (a) Complete the sample space diagram.

(2 marks)

Select the correct answer.

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

<input type="checkbox"/>	5-sided spinner					
		1	2	3	4	5
	1	2	3	4	5	6
2-sided spinner	2	3	4	5	6	7

<input type="checkbox"/>	5-sided spinner					
		1	2	3	4	5
	1	2	3	3	4	5
2-sided spinner	2	3	4	6	8	10

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

(2 marks)

Find all the numbers in the table that are 5 or larger

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

$$\text{probability} = \frac{\text{5 or larger}}{\text{total outcomes}}$$

**13** Anna organises two different cooking workshops, Workshop X and Workshop Y, to help people learn to bake cakes.

She wants to compare the two different workshops to see which is more effective.

The table shows number of participants who passed and failed the baking test for each of the two workshops.

	Passed	Failed	Total
Workshop X	40	8	48
Workshop Y	5	25	30

- (i) Find the relative risk of failing the baking test having been in Workshop X compared to Workshop Y.  
(ii) Give an interpretation of your answer to part (i).

(4 marks)

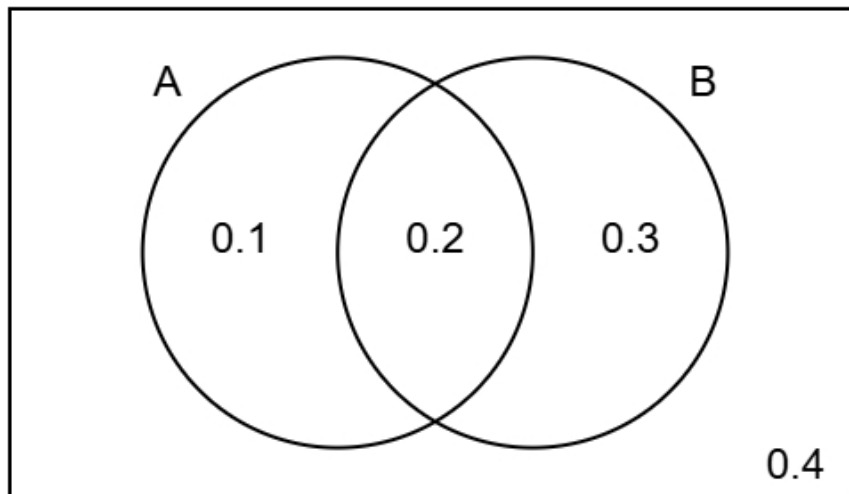
Write your answer as a decimal.

\_\_\_\_\_

Select **one** box.

- More people failed the baking test in Workshop X than in Workshop Y.
- The risk of failing the baking test having taken Workshop X is greater than the risk of failing the baking test having taken Workshop Y.
- The risk of failing the baking test having taken Workshop X is lower than the risk of failing the baking test having taken Workshop Y.
- Less people failed the baking test in Workshop X than in Workshop Y.

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



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

(1 mark)

Add the probabilities in the circle marked B together

Leave your answer as a decimal.

\_\_\_\_\_

- (b) Find  $P(A \text{ and } B)$

(1 mark)

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

Leave your answer as a decimal.

\_\_\_\_\_

(c) Find  $P(B | A)$

(2 marks)

Use the formula to find  $P(B | A)$

$$P(B | A) = \frac{P(A \text{ and } B)}{P(A)}$$

(d) Two different events events P and Q are independent.

$$P(P) = 0.6$$

$$P(Q) = 0.7$$

Find  $P(P \text{ and } Q)$

(2 marks)

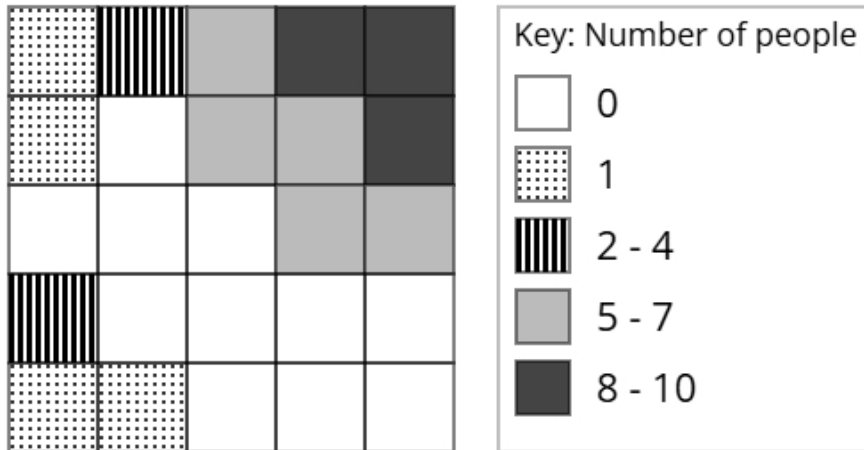
Use the formula for independent events

$$P(P \text{ and } Q) = P(P) \times P(Q)$$

Leave your answer as a decimal.

\_\_\_\_\_

- 15 The choropleth map below represents a train station concourse that has been divided into 25 squares of equal area. Mei has collected data about the popularity of different parts of the train station concourse. The number of people recorded in each square on one Sunday morning is shown.



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

(3 marks)

Find the midpoints for the groups.

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

(b) Mei would like to open a coffee kiosk in the train station concourse.

After analysing the data, she decides that she should open the coffee kiosk in the corner of the train station concourse shown at the top right of the choropleth map.

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

(2 marks)

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

because there were less people at the top right of the train station concourse.

because there were more people at the top right of the train station concourse.

Mei's comment is valid

Mei's comment is not valid

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

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

(1 mark)

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

Carol is not correct

because the data was only collected on one Sunday.

Carol is correct

because there was a large amount of data collected.