Scaffolded version

Statistics GCSE

Paper 2

2025

Edexcel Foundation

Variant 1 (same as video) 1ST0/2F

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.

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 Mr. Thompson asked the students in his maths class which animal they would most like to have from the choices Dog, Cat, Rabbit and Parrot.
Their answers are shown below:

Cat Cat Dog Cat Parrot Dog
Dog Cat Rabbit Parrot Dog Cat Dog
Dog Dog Rabbit Dog Parrot Cat Cat

(a) Complete the tally and frequency columns on the frequency table for the data.

(2 marks)

Select the correct answer.

☐

Animal	Tally	Frequency
Dog	𝍸 𝍸 𝍷	11
Cat	𝍸 𝍷𝍷𝍷𝍷	9
Rabbit	𝍷𝍷𝍷	3
Parrot	𝍷𝍷𝍷𝍷	4

☐

Animal	Tally	Frequency
Dog	𝍸 𝍷𝍷𝍷	8
Cat	𝍸 𝍷𝍷	7
Rabbit	𝍷𝍷	2
Parrot	𝍷𝍷𝍷	3

☐

Animal	Tally	Frequency
Dog	𝍸 𝍷𝍷𝍷𝍷	9
Cat	𝍸 𝍷𝍷𝍷	8
Rabbit	𝍷𝍷𝍷	3
Parrot	𝍷𝍷𝍷𝍷	4

☐

Animal	Tally	Frequency
Dog	𝍸 𝍷𝍷𝍷𝍷	9
Cat	𝍸 𝍷𝍷𝍷	8
Rabbit	𝍷𝍷	2
Parrot	𝍷𝍷𝍷	3

(b) Find the number of students in Mr. Thompson's maths class.

(1 mark)

Each person gave one answer, so count the number of answers.

____

(c) Compare the amount of students who chose Dog with the amount who chose Parrot.

(1 mark)

*Select **one** box.*

☐ The students selected Parrot and Dog in equal amounts.

☐ More students selected Dog than Parrot.

☐ More students selected Parrot than Dog.

(d) Mr. Thompson finds the mode from the data.
Give a reason why the mode is an appropriate average to use in this situation.

(1 mark)

*Select **one** box.*

☐ The data is continuous.

☐ The data is numeric.

☐ The data is not numeric.

☐ The mode will be quicker to find.

(e) Give **one** advantage of displaying this data in a tally chart rather than leaving it as raw data.

(1 mark)

*Select **one** box.*

- ☐ It is easier to analyse the data.
- ☐ It will reduce mistakes with the data.
- ☐ It improves the accuracy of the data.
- ☐ It is quicker to display the data.

2 The table shows the results of a survey into Environmental Actions and Awareness. For example, 60% of all respondents said "Recycled regularly" but only 41% of 16-24 year olds said this.
Some respondents participated in more than one option.

Environmental Action	**All People %**	**Gender**		**Age**			
		Male	**Female**	**18-26**	**27-36**	**37-56**	**57+**
Recycled regularly	60%	58%	62%	41%	52%	72%	75%
Joined a clean-up event	9%	7%	11%	5%	8%	9%	14%
Used reusable shopping bags	95%	96%	94%	99%	95%	92%	94%
Avoided single-use plastics	4%	3%	5%	6%	5%	2%	3%
Used public transport	98%	97%	99%	99%	97%	98%	98%
Bought second-hand items	82%	80%	84%	89%	85%	82%	72%
Composted food waste	35%	32%	38%	28%	31%	35%	46%
Installed energy-efficient appliances	12%	15%	9%	13%	11%	9%	15%
None of the above	1%	2%	1%	2%	1%	1%	1%

(a) A person who selected **Bought second-hand items** is chosen at random.
Describe the most likely gender and age of the person.

(2 marks)

Gender

*Select **one** box.*

☐ Female

☐ Male

Age group

Select the correct boxes.

☐ 27-36

☐ 18-26

☐ 37-56

☐ 57+

(b) Describe the relationship in the data between joining a clean-up event and age.

(1 mark)

*Select **one** box.*

☐ The proportion of people joining a clean-up event is the same at all ages.

☐ The proportion of people joining a clean-up event decreases with age.

☐ The proportion of people joining a clean-up event is highest at 57+.

☐ The proportion of people joining a clean-up event increases with age.

(c) Olivia is starting a second-hand shop.
She suspects that over half of her customers will be female.
Explain how she used the information from the table to come to this conclusion.

(1 mark)

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

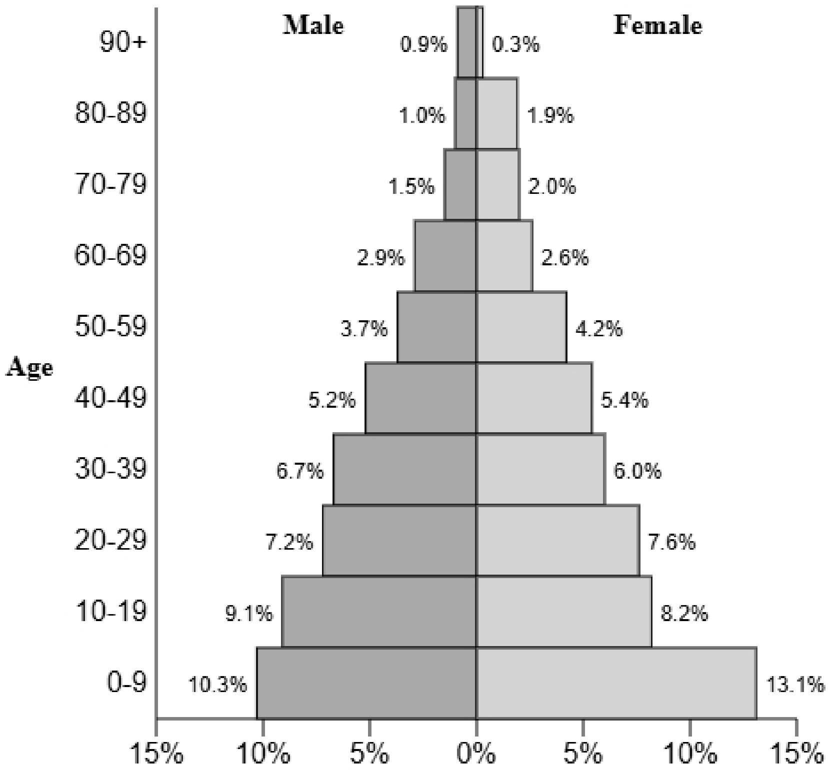
☐ but the percentage for males was higher at 84%

☐ The percentage of females who bought second-hand items was 84%

☐ but the percentage for males was lower at 80%

☐ The percentage of females who bought second-hand items was 80%

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 cinema sells four different ticket types at different prices.
The tickets for a weekend have been recorded in this table.

Ticket price	£4	£6	£8	£10	Total
Amount sold	608	400	352	240	1600

The owner of the cinema wants to open up a new cinema in a different city.
She uses the data to predict how the tickets will sell in the new location.

Ticket price	£4	£6	£8	£10
Predicted proportion	38%	25%	22%	15%

(a) The owner want to improve her predictions.
Explain how she can do this.

(1 mark)

*Select **one** box.*

☐ Draw a pie chart from the data.

☐ Rely more on the tickets she feels are popular.

☐ Only collect data on one of the days.

☐ Collect more data.

(b) The owner opens up the new cinema.
The table below shows the tickets sales over the opening weekend at the new cinema.

Ticket price	£4	£6	£8	£10
True proportion	16%	21%	28%	35%

Compare the sales of £10 tickets at the new location compared to the original location and suggest a reason for any difference.

(2 marks)

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

☐ There was a higher amount of £10 tickets at the new location.

☐ There was a higher proportion of £10 tickets at the new location.

☐ This could have been because at an opening weekend people may be happier to spend more on a ticket.

☐ This could have been because the new location was more popular.

5 Marco owns a food truck that sells pizza.
Customers can choose either cheese or pepperoni pizza **and** either a cola or a lemonade drink.
He records his sales in a two-way table, but the table is incomplete.

	cola	lemonade	Total
cheese pizza		35	80
pepperoni pizza	48	47	
Total	93		175

(a) Complete the two way table.

(2 marks)

	cola	lemonade	Total
cheese		35	80
pepperoni	48	47	
Total	93		175

(b) Find the probability a customer chosen at random:

i) ordered a cheese pizza.

ii) ordered a pepperoni pizza and a cola.

iii) did **not** order a cola.

(4 marks)

You can use the formula:

$$\text{Probability} = \frac{\text{Desired frequency}}{\text{Total frequency}}$$

The total frequency is found at the bottom-right of the table.

For part iii, we are looking at **not** lemonade, which means they must have picked cola.

Probability of cheese pizza.

Probability of pepperoni pizza and a cola.

Probability of **not** a cola.

(c) Marco is ordering drinks for the following month.
Explain, with a reason, if Marco should order more bottles of cola than lemonade.

(2 marks)

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

- ☐ Yes, Marco should buy more bottles of cola than lemonade
- ☐ No, Marco should buy less bottles of lemonade than cola
- ☐ because he sold more cola than lemonade.
- ☐ because he sold less lemonade than cola.

6 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) Give two reasons why Daniel might think this.

(2 marks)

*Select **two** boxes.*

- ☐ A sample is cheaper.
- ☐ Daniel will be able to explain each question.
- ☐ A sample is quicker.
- ☐ Daniel will be able to choose who is in the sample.
- ☐ A sample is more accurate.

(b) Daniel has decided to use the class registers as a sampling frame.
Explain what a sampling frame is.

(1 mark)

*Select **one** box.*

- ☐ A list of all the members in the population.
- ☐ The whole group.
- ☐ A list of all the members in the sample.
- ☐ The tally chart or table used in the survey.

(c) 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.

(d) Give **two** reasons why Daniel should conduct a pilot survey.

(2 marks)

*Select **two** boxes.*

- ☐ A pilot survey will help identify problems.
- ☐ A pilot survey will give more accurate data.
- ☐ A pilot survey will include more people.
- ☐ A pilot survey will give an idea of what the results may be.
- ☐ A pilot survey will be cheaper than a survey.

7 Sophia is an NHS trust manager investigating the working hours of junior doctors. She takes a simple random sample of 12 doctors in a hospital and asks them to report the number of hours they worked last week.
The working hours of the 8 people are listed:

36	41	39	70
40	38	42	38

Sophia believes that one of the values is an outlier.

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

(1 mark)

*Select **one** box.*

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

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

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

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

(b) Write down the value that is most likely to be an outlier and explain why you think this value is an outlier.

(2 marks)

The outlier is ___________ hours

*Select **one** box.*

☐ This value appears only once in the dataset.

☐ This value is close to the mean.

☐ This value is significantly higher than the other numbers.

☐ This value doesn't match the median.

(c) Sophia removes the outlier.
State whether the mean of the remaining seven values is greater than, is equal to or is less than the mean of all eight salaries.
Give a reason for your answer.

(2 marks)

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

- ☐ The mean is less.
- ☐ This is because the value that Sophia removed is higher than all the other values.
- ☐ This is because the sum of values will now be divided by seven.
- ☐ The mean is the same.
- ☐ The mean is greater.

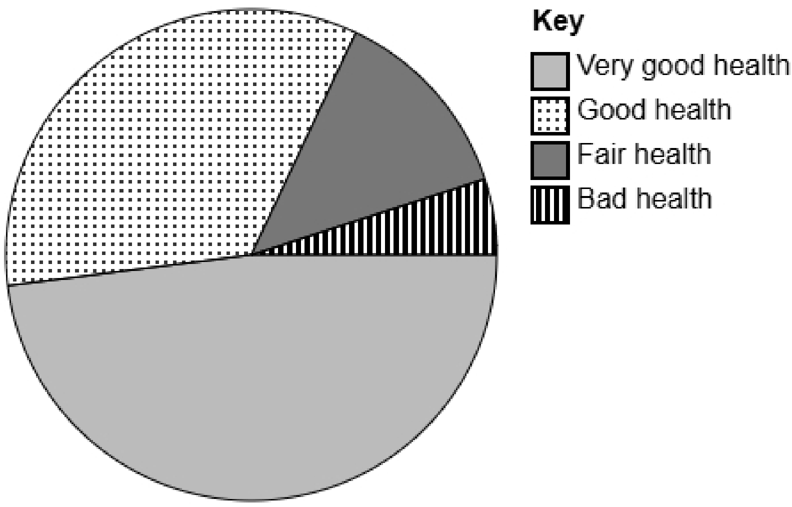
(d) After calculating the mean of the seven values without the outlier, Sophia uses this mean in a report to describe all the doctors in the NHS.
Describe two things that could affect the reliability of her conclusions.

(2 marks)

*Select **two** boxes.*

- ☐ The mean will be the same with and without the outlier.
- ☐ Removing the outlier may ignore the true variation in working hours.
- ☐ Small sample size.
- ☐ The number of hours may not truly reflect the hard work of the doctors.
- ☐ One value cannot represent many.

8 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.

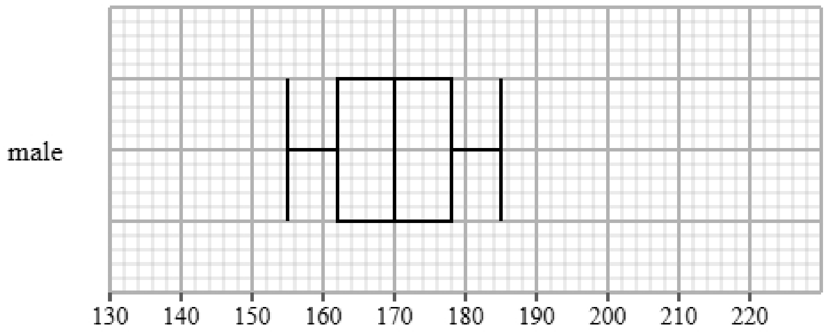
(b) The population in the UK in 2021 was estimated to be 67 million.
Calculate an estimate for the number of people in the UK in 2021 who rate themselves as having 'Good health'.
Round your answer to the nearest million.

(2 marks)

Start by measuring the angle with a protractor. Next find the amount. Remember that there are 360 degrees in a circle.

__________ million

9 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.

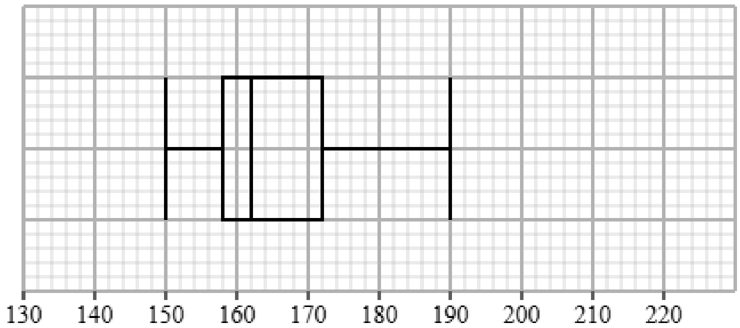
Least tall	Lower quartile	Median	Upper quartile	Most tall
150	158	162	172	190

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

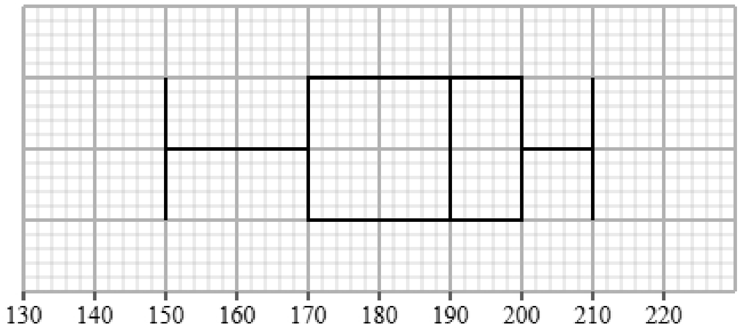
(2 marks)

Select the correct answer.

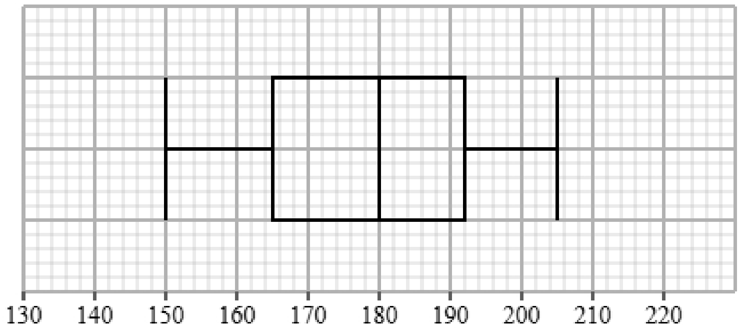
☐ female



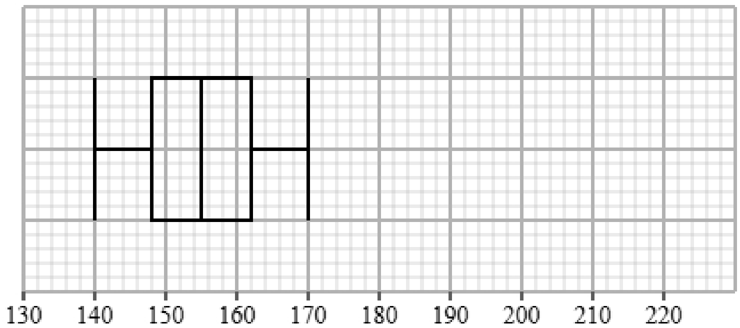
☐ female



☐ female



☐ female



(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.

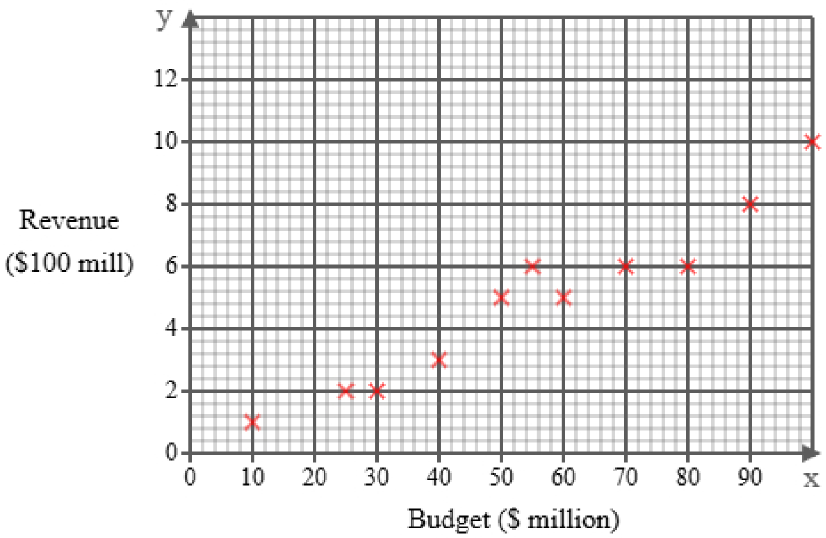
*Select **one** box.*

- ☐ 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.

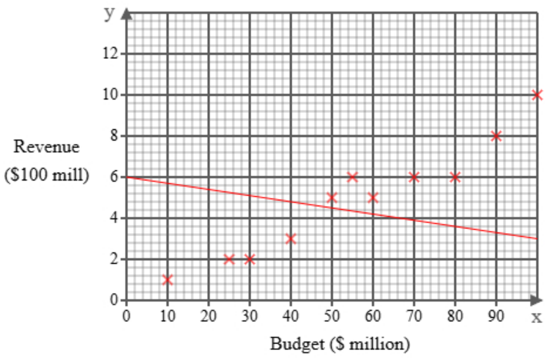
10 Liam gathered data on 11 movies, recording their production budget (in millions of dollars) and total box office revenue (in hundreds of millions of dollars). He represented his findings in the scatter diagram below.

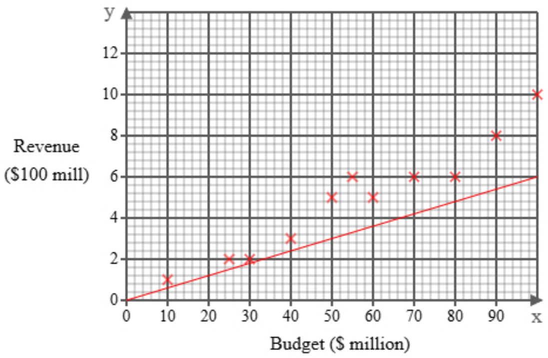


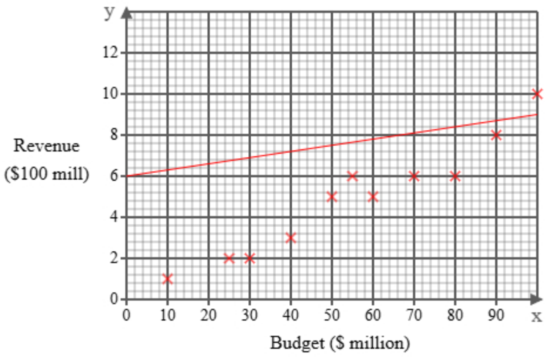
(a) Draw a line of best fit on the scatter diagram.

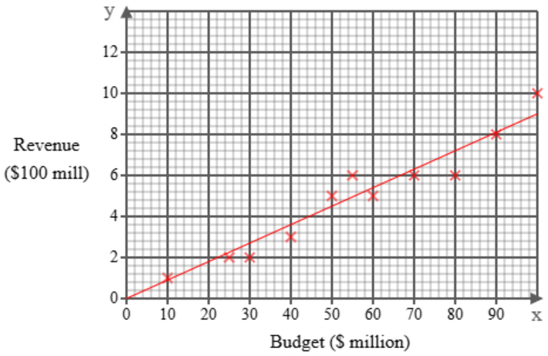
(1 mark)

Select the correct answer.

☐ 

☐ 

☐ 

☐ 

(b) 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.

(c) 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

11 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
$90 < h \leqslant 120$	0
$120 < h \leqslant 150$	37
$150 < h \leqslant 180$	147
$180 < h \leqslant 210$	26
$210 < h \leqslant 240$	0

Table B

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

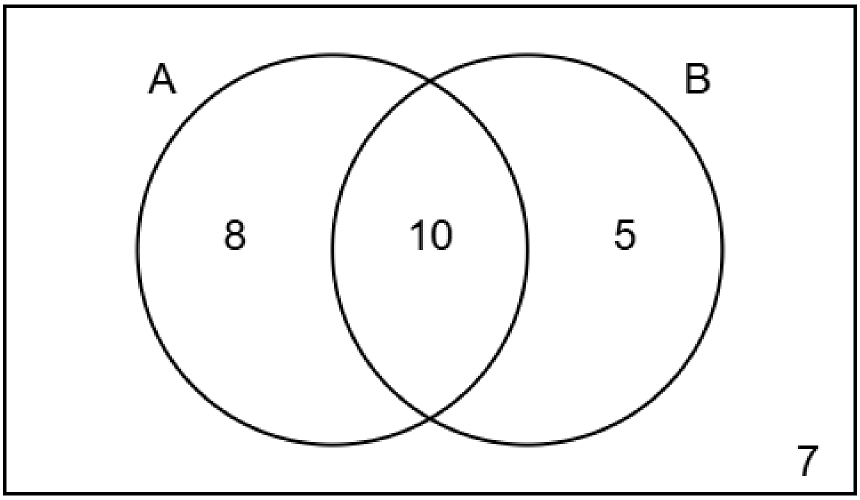
(c) Sophia wants to work out the average height of the 200 students in the school.
She decides to use Table B.
Calculate the average height of the 200 students in the school, giving your answer to 1 decimal place.

(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

12 The Venn diagram shows information about 30 books published in 2020.
A is the event that the book was written by a British author.
B is the event that the book was a bestseller (sold more than 100,000 copies).
The numbers in the Venn diagram indicate the number of books.



(a) In the Venn diagram, explain what the number 10 means.

(1 mark)

*Select **one** box.*

☐ The number of books written by a British author **or** were a bestseller

☐ The number of books written by a British author **and** were a bestseller

☐ The number of books written by a British author but are **not** a bestseller

☐ The number of books that were **not** written by a British author **or** were a bestseller

(b) Explain whether or not A and B are independent events by finding P(B) and P(B|A).

(5 marks)

Use these formulae to find P(B) and P(B|A):

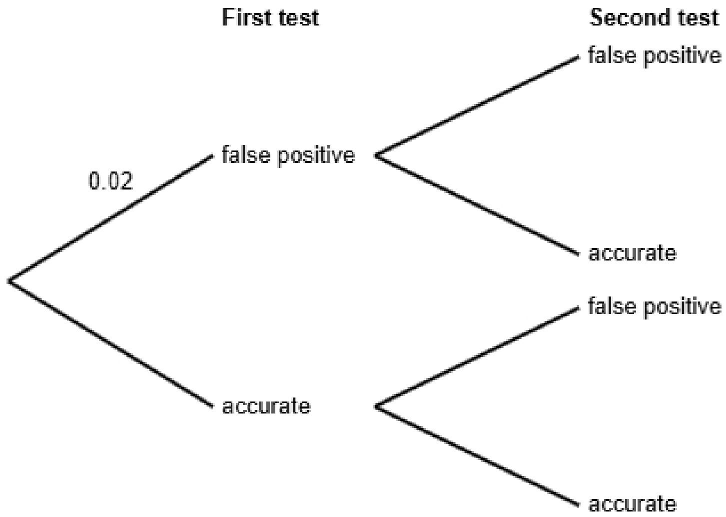
$$P(B) = \frac{\text{amount in B}}{\text{total amount}}$$

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

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

- ☐ so they are not independent
- ☐ P(B) and P(B | A) are not equal
- ☐ so they are independent
- ☐ P(B) x P(B | A) = 1

13 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)

The branches for each stage must add up to 1.
Each test is independent so will have the same probabilities.

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

(2 marks)

You will need to find P(accurate) AND P(accurate). Remember, AND means × in probability.

____

(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)

Find the probability of exactly one of the tests is false positive (there are two outcomes on the tree diagram).

____

____

____

____

*Select **one** box.*

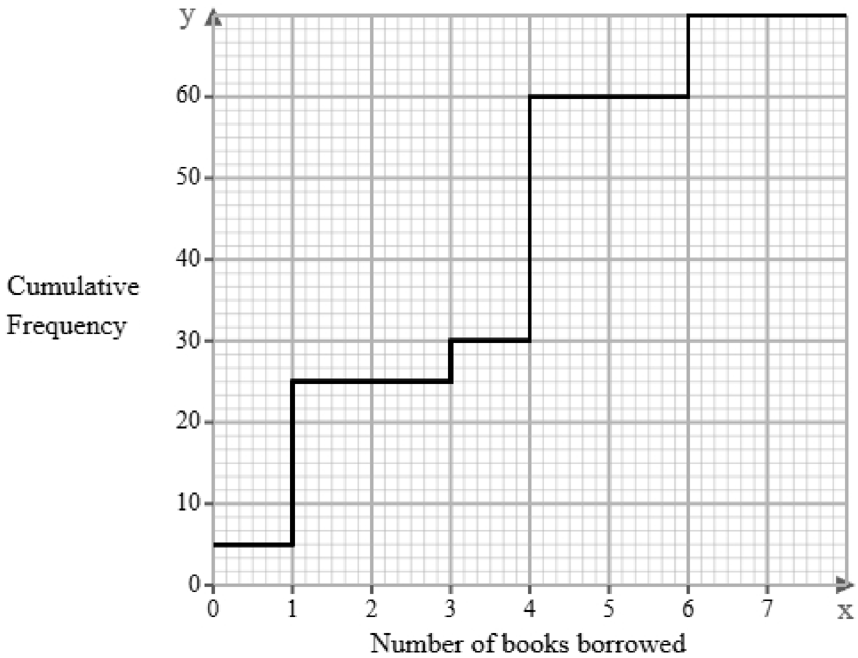
☐ The probability that exactly one of the tests is false positive is more than 4%, so David is not correct.

☐ 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.

14 The cumulative frequency step polygon shows information about the number of books borrowed from a local library over 70 days.



(a) Give a reason why a cumulative frequency step polygon has been used to display this data.

(1 mark)

*Select **one** box.*

☐ Because the number of books borrowed from a local library is continuous.

☐ Because the number of books borrowed from a local library is discrete.

☐ Because the number of books borrowed from a local library is qualitative.

☐ Because the number of books borrowed from a local library is quantitative.

(b) Find the number of days where there were:
i) exactly 2 books borrowed.
ii) more than 2 books borrowed.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point. Remember, the overall frequency is 70.

i) Exactly 2 books borrowed: __________

ii) More than 2 books borrowed: __________

(c) In 30 days fewer than x books were borrowed.
Find the value of x

(1 mark)

Draw a line across from 30 on the graph and see where all the 'jumps up' to this line are under.

____

(d) Laura believes the interquartile range of the number of books borrowed is 8.
Explain why the interquartile range for this data cannot be 8.

(1 mark)

*Select **one** box.*

- ☐ The range is 6, so the IQR must be more than 6.
- ☐ The range is 6, so the IQR must be less than 6.
- ☐ The range is 7, so the IQR must be less than 7.
- ☐ The range is 7, so the IQR must be more than 7.