statsGCSE

Scaffolded version

Statistics GCSE

Paper 2

Edexcel Foundation - 2026

Foundation Tier

Variant 4 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.

Copyright

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

Sophia 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)

Impossible – This means it cannot happen at all.

Unlikely – This means it could happen, but it probably won't.

Evens – This means it has a 50/50 chance of happening – it's just as likely to happen as not.

Likely – This means it will probably happen, but it's not guaranteed.

Certain – This means it will definitely happen – there's no doubt at all.

*Select **one** box.*

- ☐ evens
- ☐ likely
- ☐ unlikely
- ☐ impossible

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

(1 mark)

Impossible – This means it cannot happen at all.

Unlikely – This means it could happen, but it probably won't.

Evens – This means it has a 50/50 chance of happening – it's just as likely to happen as not.

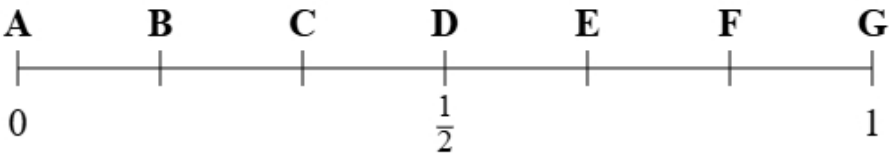
Likely – This means it will probably happen, but it's not guaranteed.

Certain – This means it will definitely happen – there's no doubt at all.

*Select **one** box.*

- ☐ likely
- ☐ evens
- ☐ certain
- ☐ unlikely

(c)



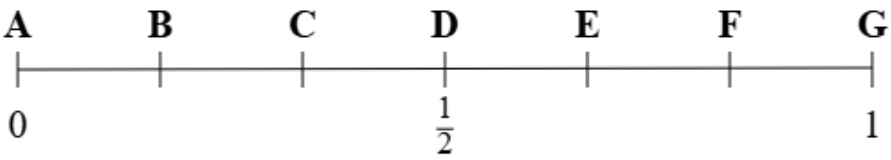
Using the probability scale, write down the letter that shows probability that the tile has a 2 on it.

(1 mark)

*Select **one** box.*

- ☐ A
- ☐ C
- ☐ B
- ☐ D

(d)



Using the probability scale, write down the letter that shows probability that the tile has a 1 or a 3 on it.

(1 mark)

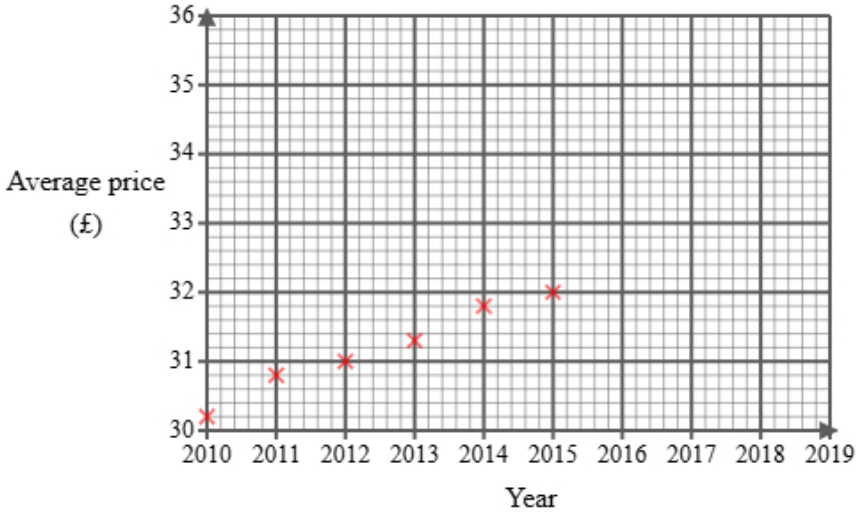
*Select **one** box.*

- ☐ F
- ☐ E
- ☐ G
- ☐ D

2 Maya found the following information about the average price of a train ticket in England.

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Average price (£)	30.20	30.80	31.00	31.30	31.80	32.00		32.80	33.00	33.50

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



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

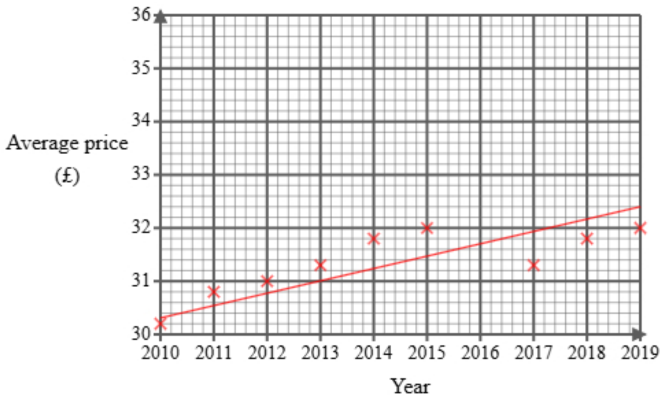
(a) (i) Plot the average price for each of 2017, 2018 and 2019

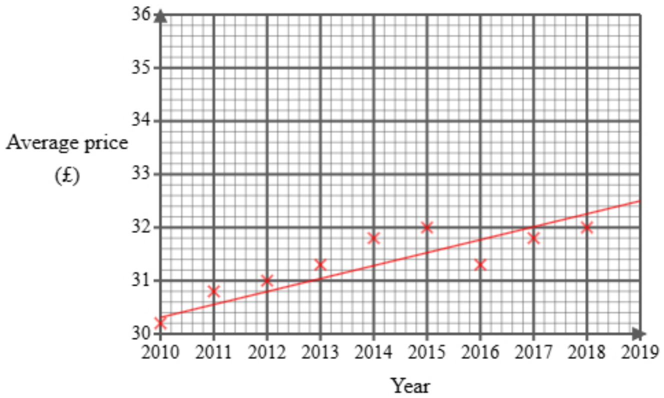
(ii) Draw a trend line for Maya's data

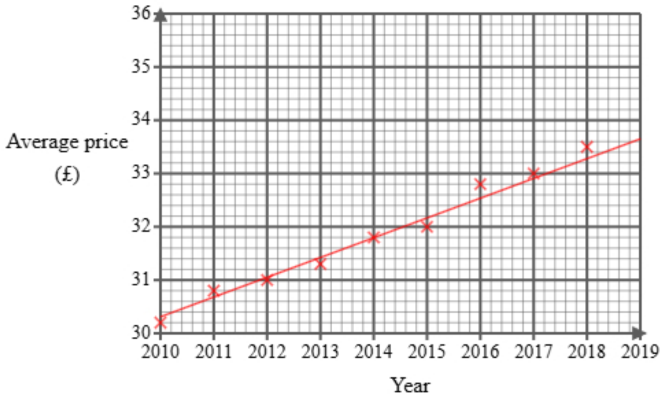
(iii) Describe the trend in the average price of a train ticket in the UK from 2010 to 2019

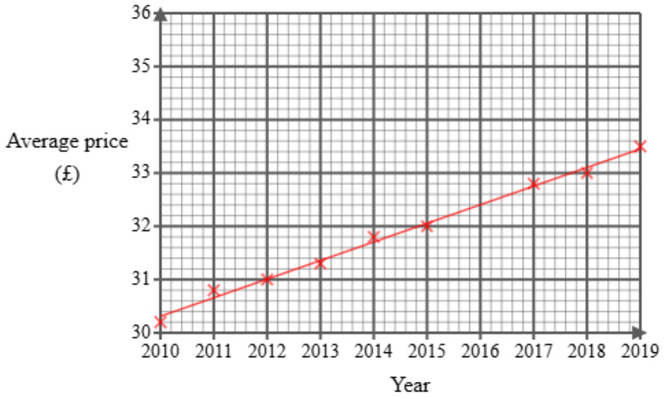
(4 marks)

Select the correct answer.

- [ ] 

- [ ] 

- [ ] 

- [ ] 

Select the correct boxes.

- ☐ Positive correlation
- ☐ Negative correlation
- ☐ Increasing
- ☐ Decreasing

(b) The gradient of Maya's trend line is 0.35

Interpret this gradient.

(1 mark)

*Select **one** box.*

- ☐ The average price increases.
- ☐ The average price started at 35 pence in 2010.
- ☐ The average price increases per year.
- ☐ The average price decreases.

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

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

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

3 A clothing store tracks how many pairs of jeans are sold each day. This information helps with restocking and running special sales based on demand.

Monday	
Tuesday	
Wednesday	
Thursday	
Friday	

Key:

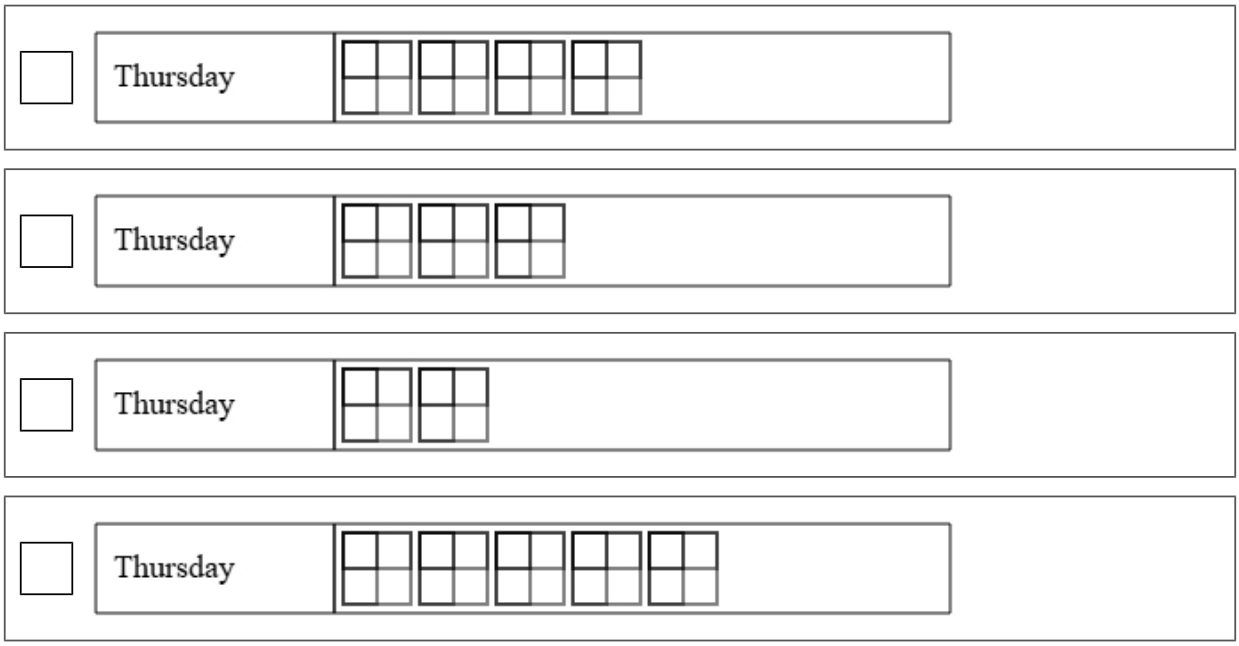
 represents 8 pairs of jeans

(a) On Thursday, the number of pairs of jeans sold was 16.

Show this information on the pictogram.

(1 mark)

Select the correct answer.



(b) Halle suggests redrawing the pictogram using a key with a whole-square representing 5 pairs of jeans.

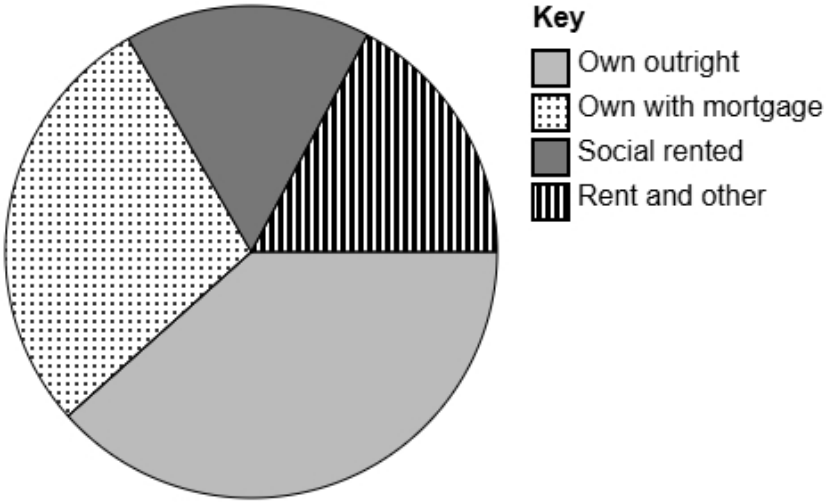
Explain why this key would **not** be suitable.

(1 mark)

*Select **one** box.*

- ☐ Tuesday shows 32 loaves of bread. This would be very difficult to show because 32 has a remainder 2 when divided by 5.
- ☐ This would be much better because you can fit more squares on.
- ☐ The key must always be an even number.
- ☐ If Halle uses 5 pairs of jeans for the key this will lead to less sales being shown.

4 The accurately drawn pie chart shows information about the tenure types for people in Wales in 2021.



(a) Explain how you can tell that most households own their house outright in Wales in 2021 using the pie chart.

(1 mark)

*Select **one** box.*

☐ 'Own outright' has the largest sector.

☐ 'Own outright' is the first value in the key.

☐ 'Own outright' is the most positive response.

☐ 'Own outright' is at the bottom of the pie chart.

(b) In 2021 there was an estimated 130 thousand households.

Calculate an estimate for the number of people in the UK in 2021 who's highest level of qualification was 'Own outright'.

Round your answer to the nearest thousand.

(2 marks)

Start by measuring the angle with a protractor.

Next find the amount. Remember that there are 360 degrees in a circle.

_____ thousand

5 Alex owns a music shop.

He wants to collect information about types of music liked by people in his town.

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

music genre

average song length

number of band members

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

(1 mark)

Categorical data can be grouped into non-overlapping categories.

*Select **one** box.*

- ☐ music genre
- ☐ number of band members
- ☐ average song length

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

(1 mark)

Discrete data can take one of a set of certain values.

*Select **one** box.*

- ☐ music genre
- ☐ average song length
- ☐ number of band members

(c) Alex would like to send a questionnaire to 30 of his customers.

He has a list of all 90 of his customers.

Explain how Alex can select a systematic sample of 30 people from his 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.
- ☐ Calculate a random number between 1 and 3 to use a starting point.
- ☐ Select every 3rd person.
- ☐ Select every 4th person.

6 Liam is a student and wants to study favourite school subjects.

He would like to find out the most liked subject in his school.

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

School subject	Tally
Maths	
English	
Science	
History	

(a) State the population for this study.

(1 mark)

*Select **one** box.*

- ☐ All students in the UK
- ☐ The students he asks
- ☐ A selection of students in Liam's school
- ☐ All the students in Liam's school

(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 people who are easiest to reach
- ☐ Sampling people randomly
- ☐ Sampling every nth person from a list

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

(1 mark)

*Select **one** box.*

- ☐ It may be biased
- ☐ It is expensive to do
- ☐ It requires a large population
- ☐ It takes too much time

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

(2 marks)

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

- ☐ It will make it easy for Liam to analyse the data.
- ☐ Liam may not ask enough students.
- ☐ There is no 'other' option.
- ☐ It will make collecting data very difficult.

(e) After collecting the data, Liam 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.
- ☐ It is suitable
- ☐ It is not suitable
- ☐ because the data is qualitative.

7 A trainer wants to get feedback on a workplace training session they ran last week.

41 people attended the training session.

The trainer plans to give a questionnaire to a sample of 10 of the people who attended the training session.

One of the questions on the questionnaire is:

> To what extent do you agree with the statement, the trainer was engaging?
>
> 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.*

- ☐ ordinal
- ☐ continuous
- ☐ bivariate

(b) Explain how the trainer could use a list of random numbers to choose a simple random sample of 10 people who attended the training session.

(3 marks)

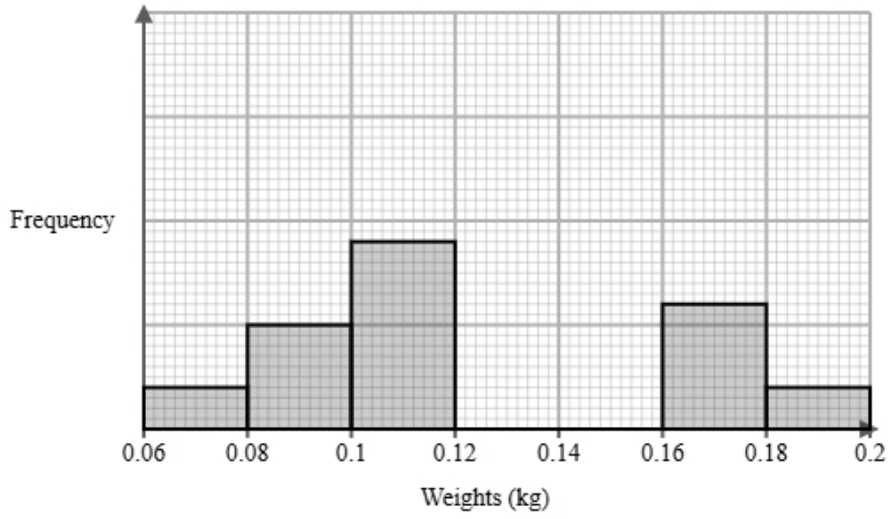
*Select the **three** correct statements (**three** statements are incorrect).*

- ☐ Select 10 people from a hat.
- ☐ Select the people who were assigned the highest number.
- ☐ Select 10 numbers ignoring any numbers that have repeated or are out of range.
- ☐ Assign a number to all the people.
- ☐ List the people in alphabetical order based on their surname.
- ☐ Select people with the corresponding number from random number list.

8 A farmer measured the weights of King Edward potatoes and sweet potatoes in their field.

They recorded the weights after 4 months.

The incomplete histogram and grouped frequency table give information about the weights of King Edward potatoes in their field.



Weights w (kg)	Frequency
$0.06 < w \leqslant 0.08$	2
$0.08 < w \leqslant 0.10$	5
$0.10 < w \leqslant 0.12$	9
$0.12 < w \leqslant 0.14$	12
$0.14 < w \leqslant 0.16$	9
$0.16 < w \leqslant 0.18$	
$0.18 < w \leqslant 0.2$	

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

(2 marks)

Find the height of $0.06 < w \leqslant 0.08$ on the graph and compare it with the frequency to find the scale

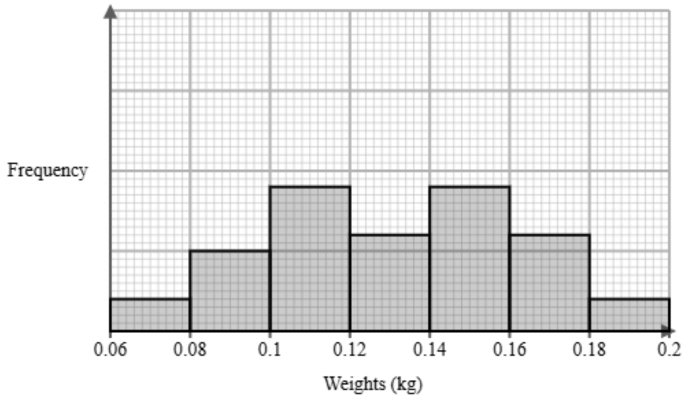
Use this scale to find the missing frequencies

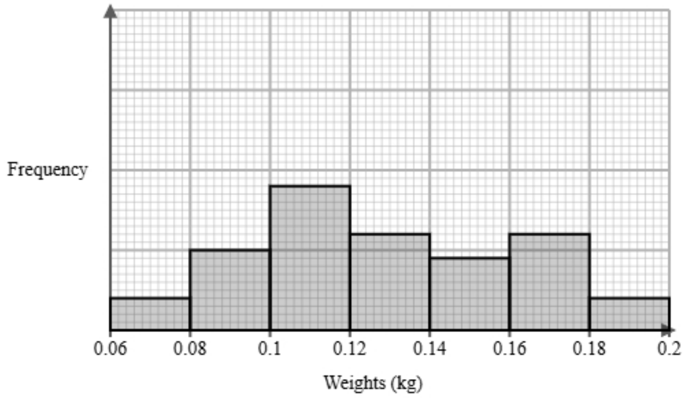
Weights w (kg)	Frequency
$0.06 < w \leqslant 0.08$	2
$0.08 < w \leqslant 0.10$	5
$0.10 < w \leqslant 0.12$	9
$0.12 < w \leqslant 0.14$	12
$0.14 < w \leqslant 0.16$	9
$0.16 < w \leqslant 0.18$	
$0.18 < w \leqslant 0.2$	

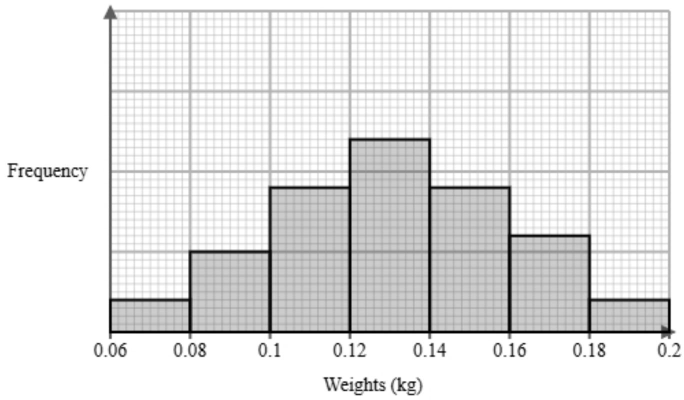
(b) Use the information in the table to complete the histogram.

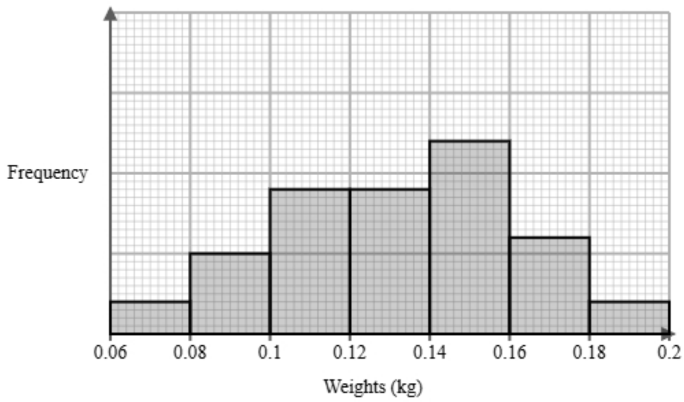
(2 marks)

Select the correct answer.

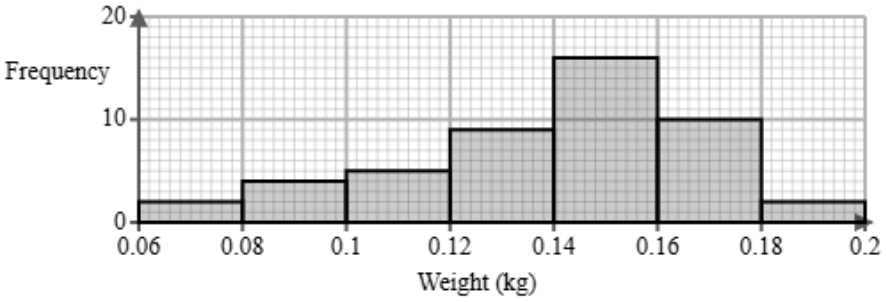
- [ ] 

- [ ] 

- [ ] 

- [ ] 

(c) The histogram below shows data on the weights of sweet potatoes after 4 months.



Identify and interpret the type of skew shown in the histogram for sweet potatoes.

(2 marks)

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

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

(d) The farmer also measured the weights of King Edward potatoes and sweet potatoes in a different field.
The grouped frequency table below gives information about the weights of King Edward potatoes and sweet potatoes in a different field.

	Frequency	
Weights w (grams)	King Edward potatoes	sweet potatoes
$80 < w \leqslant 100$	2	1
$100 < w \leqslant 120$	18	8
$120 < w \leqslant 140$	17	16
$140 < w \leqslant 160$	8	9
Total	45	34

The estimate of the mean for King Edward potatoes is calculated to be 123.8 g to 1 decimal place.
Rajesh uses the estimate of the means for King Edward potatoes and sweet potatoes to conclude that the King Edward potatoes weigh more than sweet potatoes.

Discuss whether or not Rajesh is correct and give **one** limitation of your conclusion.
You must show your working.

(4 marks)

Add a midpoint and fw column onto the table
Add up the frequencies and the fw columns
Find the mean by dividing $\sum fw$ and $\sum f$

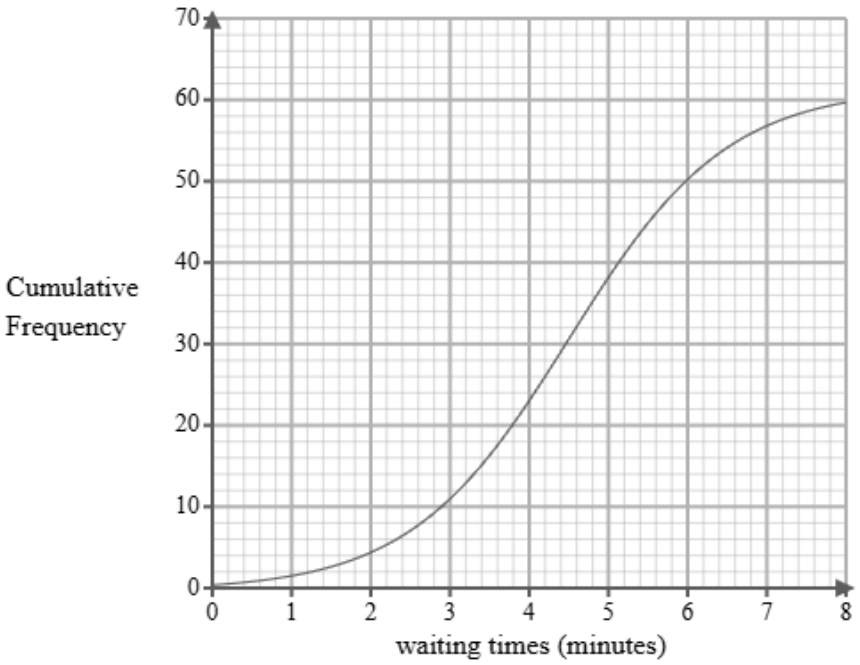
Mean of sweet potatoes = _____

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

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

9 A researcher measures the waiting times, in minutes, of 60 customers at a café.

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.

☐

Lower quartile	Median	Upper quartile
2.1	5	7.2

☐

Lower quartile	Median	Upper quartile
2.9	5	6.3

☐

Lower quartile	Median	Upper quartile
3.4	4.5	5.5

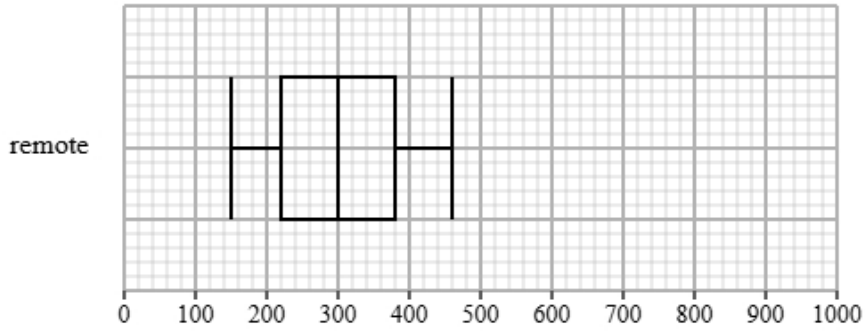
☐

Lower quartile	Median	Upper quartile
3.8	4.5	4.7

10 Ethan collected the steps for remote and on-site workers in an hour within their day.

Both groups recorded their steps over the same period.

The box plot presents data on the steps for the remote workers.



The table gives information about the steps for the on-site workers.

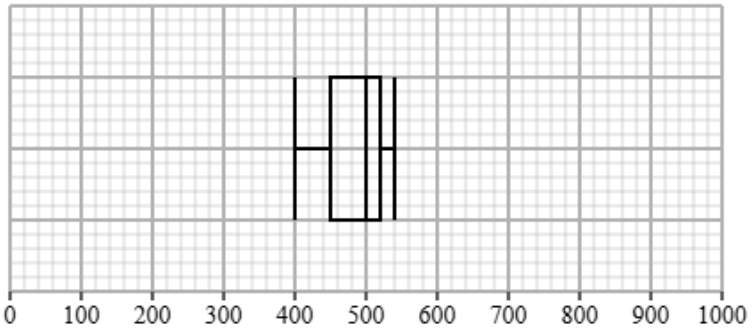
Least tall	Lower quartile	Median	Upper quartile	Most tall
300	550	800	850	900

(a) Draw a box plot for the steps for the on-site workers.

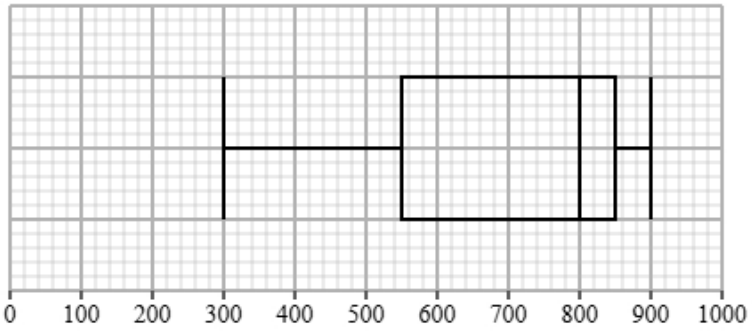
(2 marks)

Select the correct answer.

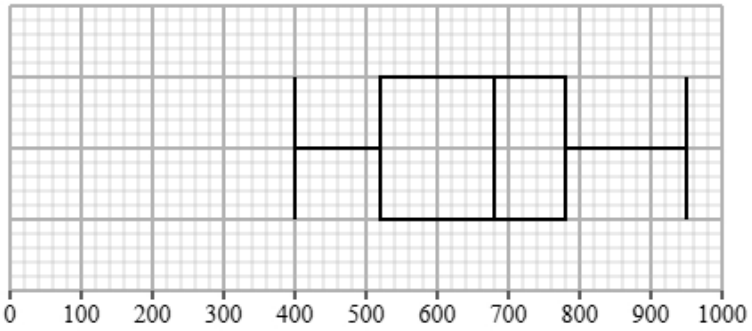
☐ on-site



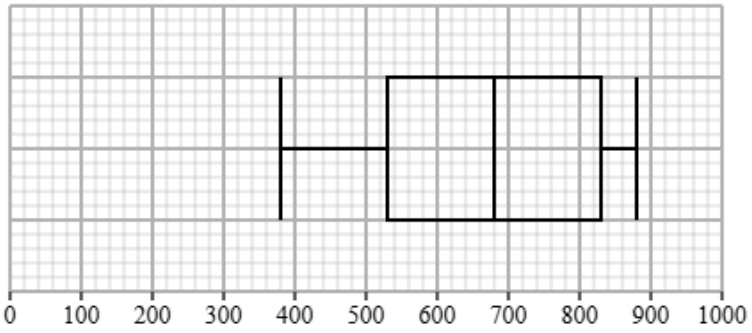
☐ on-site



☐ on-site



☐ on-site



(b) Compare the two distributions of steps.

Give three comparisons and interpret one of these comparisons.

(4 marks)

*Select **one** box.*

- ☐ The median is bigger.
- ☐ The median steps for remote workers is greater than on-site workers.
- ☐ The median steps for remote workers is lower than on-site workers.
- ☐ The median steps for remote and on-site workers are equal.

*Select **one** box.*

- ☐ The IQR is bigger.
- ☐ The IQR for the steps of the remote workers is greater than on-site workers.
- ☐ The IQR for the steps of the remote workers is lower than on-site workers.
- ☐ The IQR for the steps of the remote and on-site workers are equal.

*Select **one** box.*

- ☐ The skews for the steps of the remote and on-site workers are both positive.
- ☐ The skew for the steps of the remote workers is symmetrical and the skew for the on-site workers is positive.
- ☐ The skew for the steps of the remote workers is symmetrical and the skew for the on-site workers is negative.
- ☐ The skews for the steps of the remote and on-site workers are both symmetrical.

*Select **one** box.*

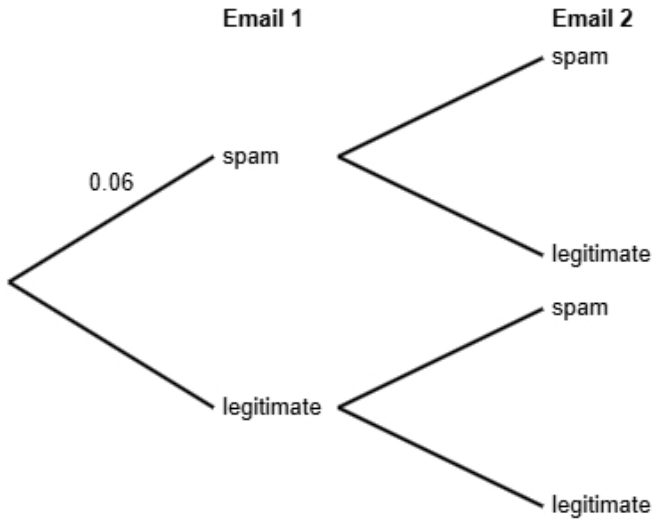
- ☐ The steps for the remote workers are more spread out than the on-site workers.
- ☐ The remote workers on average walk less than the on-site workers.
- ☐ The remote workers on average walk more than the on-site workers.
- ☐ The remote workers are more skewed than on-site workers.

11 A study shows that 6% of emails received by a certain email provider are spam.

All other emails are legitimate.

Maria receives two emails in her inbox.

She does not know if each email is spam or legitimate.



(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 Maria's emails are legitimate.

(2 marks)

You will need to find P(legitimate) AND P(legitimate).

Remember, AND means × in probability.

(c) Maria states that the probability that exactly one email is spam is less than 12%

Find out whether or not Maria is correct.

(3 marks)

Find the probability of exactly one email is spam (there are two outcomes on the tree diagram).

*Select **one** box.*

- ☐ The probability that exactly one email is spam is less than 12%, so Maria is not correct.
- ☐ The probability that exactly one email is spam is less than 12%, so Maria is correct.
- ☐ The probability that exactly one email is spam is more than 12%, so Maria is correct.
- ☐ The probability that exactly one email is spam is more than 12%, so Maria is not correct.

12 The table shows information about bicycles for sale in Birmingham.

gear types	number of bicycles
1	175
2	350
3	200
4	450
5 or more	425
Total	1600

A researcher wants to investigate the price of these bicycles and takes a stratified sample of 64 bicycles according to the gear types.

(a) The researcher says the mode of the gear types for these bicycles is 4.

Explain how the researcher knows this.

(1 mark)

*Select **one** box.*

☐ 4 gears has the highest frequency.

☐ 4 bicycles has the highest frequency.

☐ 4 is the difference between the largest and smallest number.

☐ 4 is the middle number.

(b) Work out the number of bicycles in the sample for each gear type.

gear types	number of bicycles in the sample
1	
2	
3	
4	
5 or more	

(3 marks)

Find the divisor for the stratified sample: $\frac{\text{total}}{\text{sample size}}$

Divide each frequency by this number to find the required sample in each group

(c) Describe how the 64 bicycles in the sample should be selected.

(3 marks)

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

- ☐ Ensure that all 1600 bicycles are included in the sample.
- ☐ Select the first 64 bicycles.
- ☐ Generate random numbers, remove repeats or numbers out of range.
- ☐ Use a sampling frame for each strata.
- ☐ Complete two of the strata.
- ☐ Number each of the bicycles, and then use the random numbers to select the required amount of bicycles.

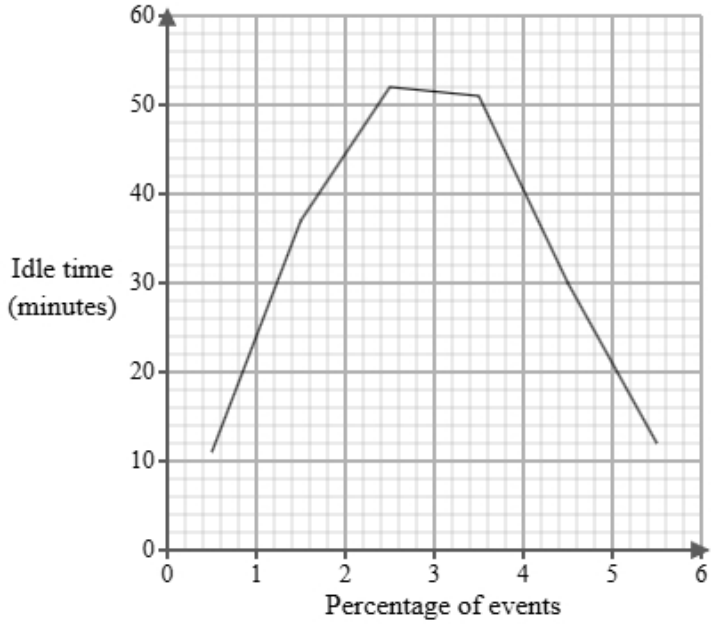
13 Tom works for a logistics company. He has been tasked with investigating delivery vehicle idle times. Below is a section of the spreadsheet he used to record his findings.

Idle time (Minutes)	Percentage of events
$0 < d \leqslant 1$	6
$1 < d \leqslant 2$	four
$2 < d \leqslant 3$	9
$3 < d \leqslant 4$	119
$4 < d \leqslant 5$	47
$5 < d \leqslant 6$	15
Total	100

Tom cleans the data to create the table below.

Idle time (Minutes)	Percentage of events
$0 < d \leqslant 1$	6
$1 < d \leqslant 2$	4
$2 < d \leqslant 3$	9
$3 < d \leqslant 4$	19
$4 < d \leqslant 5$	47
$5 < d \leqslant 6$	15
Total	100

A frequency polygon has been drawn for delivery vehicle idle times at a rival firm.



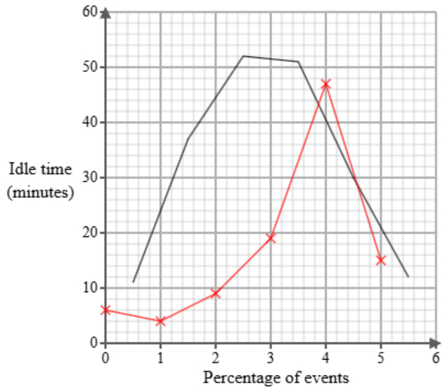
i) On the same graph, draw the frequency polygon for delivery vehicle idle times for Tom's company.

ii) Using the two frequency polygons, compare the skew of the distributions and explain what your comparison means in context.

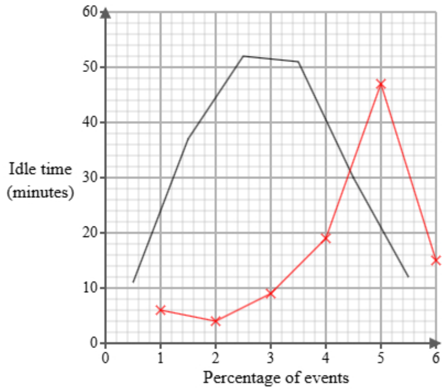
(4 marks)

Select the correct answer.

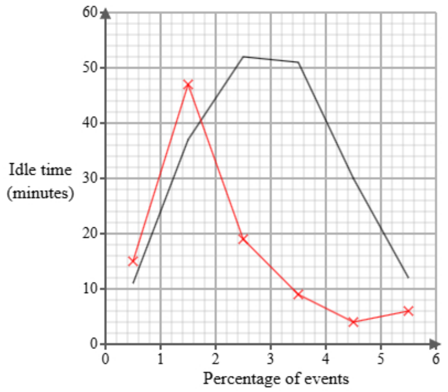
☐



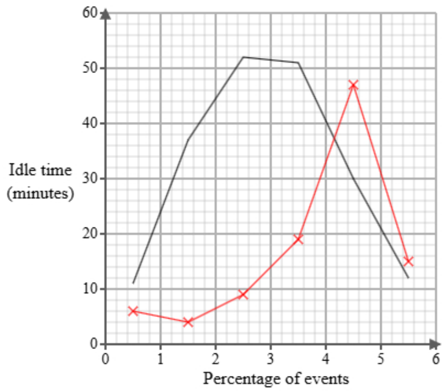
☐



☐



☐



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

- ☐ The distribution of delivery vehicle idle times at a rival firm is symmetrical whereas the distribution of delivery vehicle idle times for Tom's company is negatively skewed.
- ☐ The distribution of delivery vehicle idle times at a rival firm is positively skewed whereas the distribution of delivery vehicle idle times for Tom's company is symmetrical.
- ☐ This means that idle times for the rival firm were equally spread out on either side of the median and idle times for the for Tom's company were mainly at the upper end of the distribution.
- ☐ The distribution of delivery vehicle idle times at a rival firm is negatively skewed whereas the distribution of delivery vehicle idle times for Tom's company is symmetrical.
- ☐ This means that idle times for the rival firm were mainly at the lower end of the distribution and idle times for the for Tom's company were mainly at the upper end of the distribution.
- ☐ This means that idle times for the rival firm were mainly at the upper end of the distribution and idle times for the for Tom's company were equally spread out on either side of the median.

14 The table shows information about the retail price index (RPI) and cinema ticket price (£) in the United Kingdom for Jan 1990, Jan 2000 and Jan 2010.

	Jan 1990	**Jan 2000**	**Jan 2010**
retail price index	100	145	177
cinema ticket price (£)	2.81	3.99	5.97

Describe how the increase in cinema ticket price (£) compares with the RPI over the ten years to Jan 2000 and over the twenty years to Jan 2010.

(5 marks)

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

☐ Between Jan 1990 and Jan 2000 the change in price was less than the RPI.

☐ Between Jan 1990 and Jan 2000 the change in price was more than the RPI.

☐ $\frac{3.99}{145} \times 100 = 3$ (nearest integer)

☐ Between Jan 1990 and Jan 2010 the change in price was less than the RPI.

☐ $\frac{5.97}{2.81} \times 100 = 212$ (nearest integer)

☐ $\frac{5.97}{177} \times 100 = 3$ (nearest integer)

☐ Between Jan 1990 and Jan 2010 the change in price was more than the RPI.

☐ $\frac{3.99}{2.81} \times 100 = 142$ (nearest integer)

15 Daniel is investigating how the distance from city centre in km, x, affects the selling price (£), y for two types of houses, detached houses and semi-detached houses.

He found ten houses of each type and recorded their distance from city centre and selling price and plotted each on scatter diagrams.

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

Here are the results:

House type	Gradient of line of best fit	y-intercept of line of best fit
Detached houses	-5000	450000
Semi-detached houses	-4200	380000

Interpret and compare these results in context.

(5 marks)

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

- ☐ House type Semi-detached houses reduces in selling price by £4200 per km.
- ☐ House type Semi-detached houses has a greater initial selling price.
- ☐ House type Detached houses reduces in selling price less than House type Semi-detached houses.
- ☐ House type Detached houses reduces in selling price more per km than House type Semi-detached houses.
- ☐ Both houses increase in selling price as the distance from city centre increase.
- ☐ House type Detached houses changes in selling price by £450000 per km.
- ☐ House type Detached houses has a greater initial selling price.
- ☐ House type Detached houses reduces in selling price by £5000 per km.
- ☐ Both houses decrease in selling price as the distance from city centre increase.
- ☐ House type Semi-detached houses changes in selling price by £380000 per km.