

Statistics GCSE**Paper 2**

2025

Edexcel Higher

Variant 4

1ST0/2H

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.

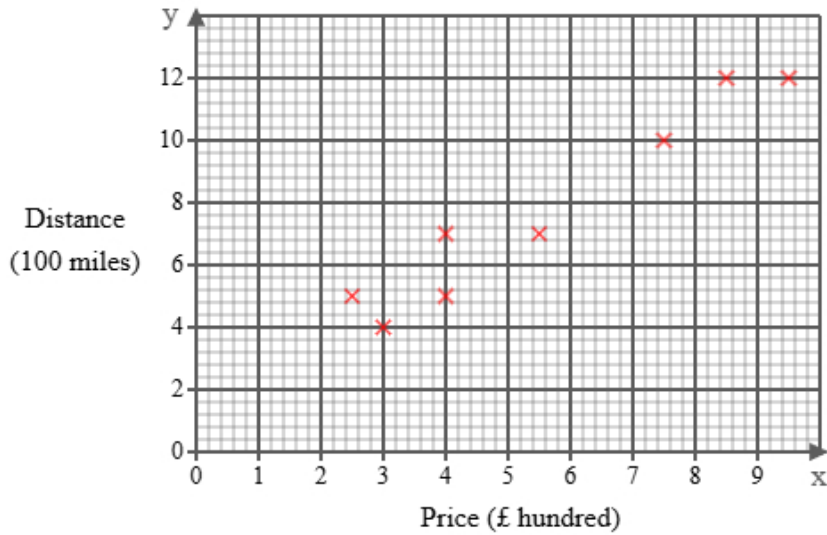
Copyright

- 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 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 price.

(1 mark)

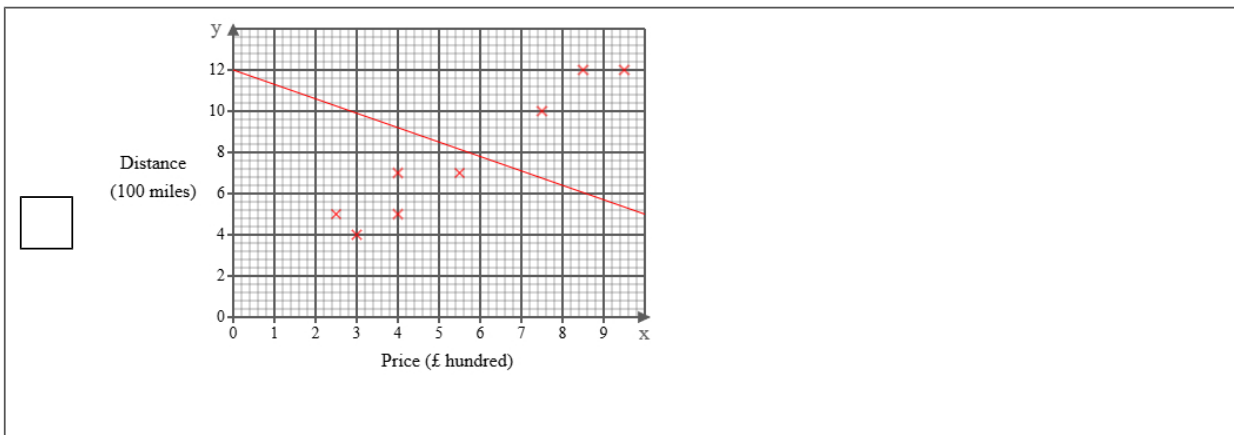
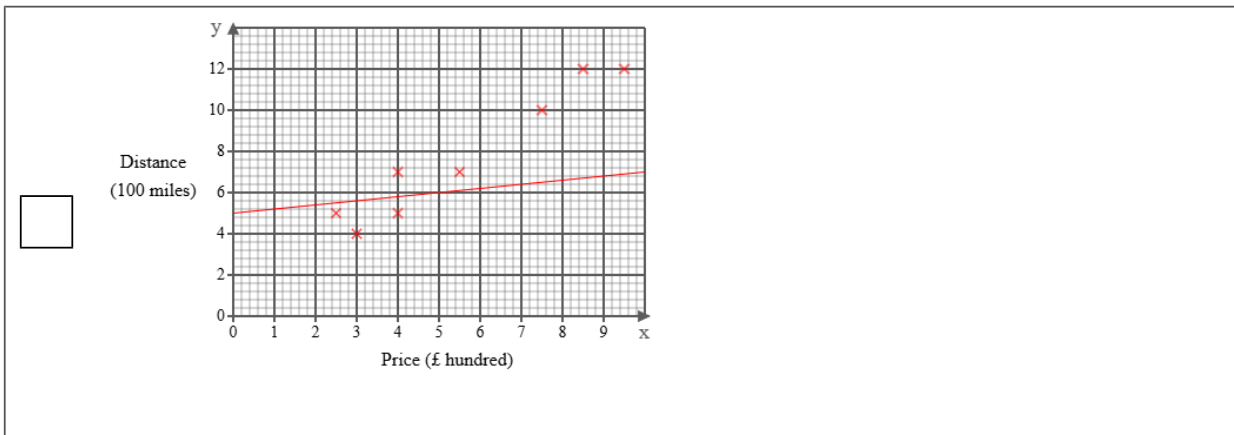
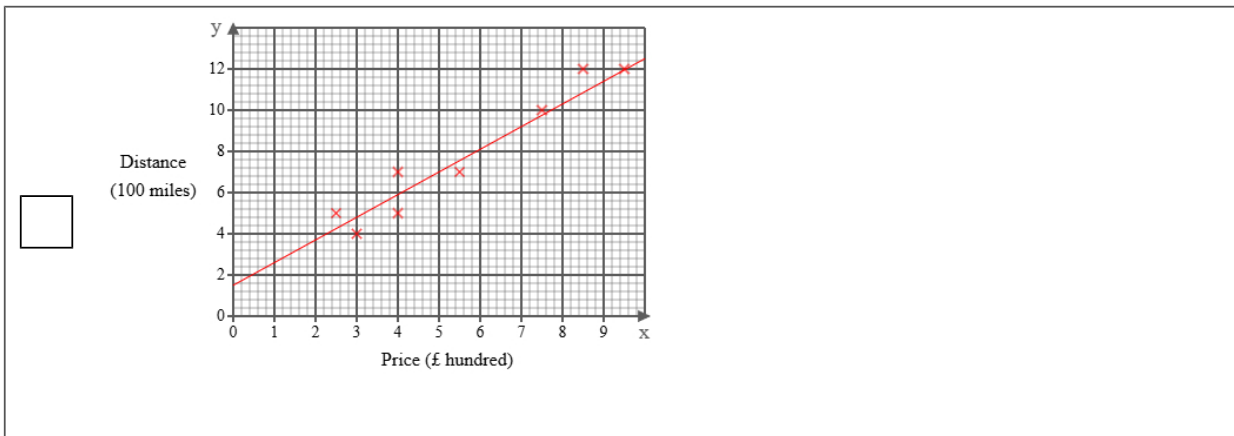
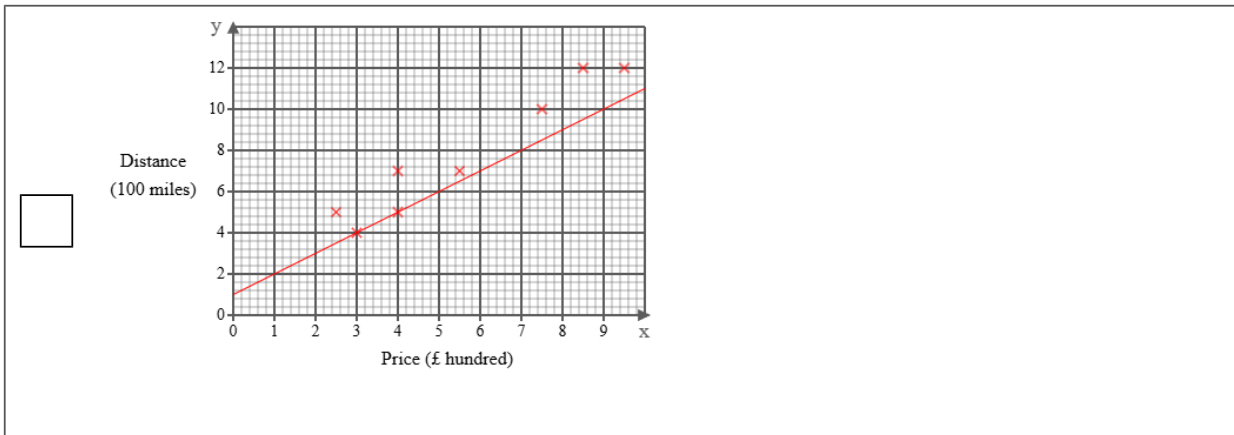
Find the cross on the scatter graph that is at [s036]550 on the x-axis (the bottom axis), then read off the value from the y-axis (the side axis).

_____ 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.

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.

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

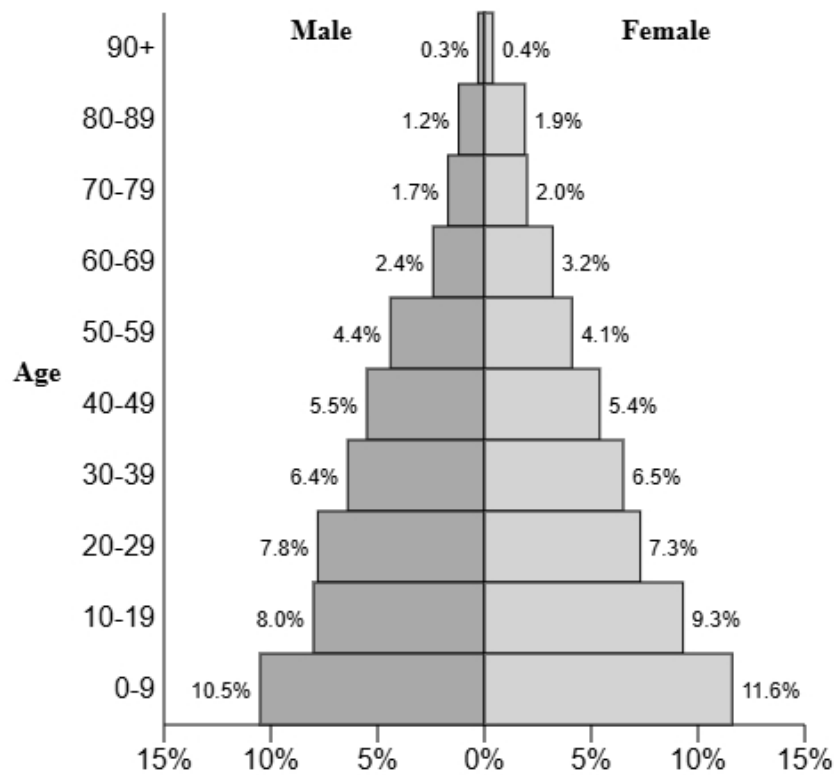
State one problem Sophia may have using the electoral register as a sampling frame.

(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.

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



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

(1 mark)

Select *one* box.

- 40-49
- 20-29
- 30-39
- 10-19

(b) 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.

(c) 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.

A scientist is conducting an experiment to investigate whether meditation improves focus.

She plans to use a matched pairs design.

Priya is one of the participants in the study.

As part of the experiment, she completes four focus-related tasks.

Each test has a different weighting.

The table below shows the weightings and Priya's scores for each test.

Test	Weighting	Score
A	4	30
B	1	12
C	2	20
D	3	27

- 4 Calculate the weighted mean score for Priya's four tests.

(3 marks)

Use the formula

$$\text{weighted mean} = \frac{\sum w \times s}{\sum w}$$

- 5 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
$0 < d \leq 1$	6
$1 < d \leq 2$	four
$2 < d \leq 3$	9
$3 < d \leq 4$	119
$4 < d \leq 5$	47
$5 < d \leq 6$	15
Total	100

Tom cleans the data to create the table below.

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

- (a) Give a reason Tom cleaned the data.

(1 mark)

Select **one** box.

- The data should be spread evenly between each group.
- One of the percentages says four not 4.
- It should show frequencies not percentages.

(b) Tom realised that the value of 119 in the original table was incorrect.
Explain how Tom knew this.

(1 mark)

Select **one** box.

- Tom removed it because it was an outlier.
- The total is 100, so no value can be more than this.
- Tom wanted the data to be closer together.

(c) Use linear interpolation to work out an estimate of the median idle time.
Round your answer to one decimal place.

(3 marks)

_____ minutes

- 6 The table shows the average monthly copper price, in euros, over a six-month period in 2020.
It also provides selected chain base index numbers, rounded to one decimal place, for the same data.

	Jan	Feb	Mar	Apr	May	Jun
copper price	5000	5100	5200	5300	5250	5400
Chain base index number		102	102	101.9	99.1	

- (a) Calculate the chain base index for June.
Round your answer to one decimal place.

(2 marks)

Use the formula:

$$\text{current index} = \frac{\text{current value}}{\text{previous value}} \times \text{previous index}$$

- (b) i) Calculate the geometric mean of the five chain base index numbers, showing all your working and rounding your answer to one decimal place.
ii) Interpret your answer in context.

(4 marks)

For part i, use the formula:

$$\sqrt[5]{102 \times 102 \times 101.9 \times 99.1 \times 102.9}$$

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

- On average, copper prices decreased
- by about 1.2% per month.
- by about 1.6% per month.
- On average, copper prices increased

7 At a comic convention , 80% of visitors are in-costume visitors and 20% are plain-clothed visitors.

Isabella and Jake plan to conduct a feedback survey.

Isabella decides to use simple random sampling to select 50 participants.

She uses the official attendee list as a sampling frame, assigning a number to each participant.

She then generates 50 random numbers and selects her sample accordingly.

Jake decides to use quota sampling to collect a sample of 50 participants.

He plans to sit one of the entrances until 40 in-costume visitors and 10 plain-clothed visitors have been interviewed.

(a) Give two reasons why Isabella's method may **not** produce a sample of 50 participants.

(2 marks)

Select **two** boxes.

- A person may want to participate but not be picked.
- The person may have joined.
- Some people may have left the hospital.
- There may not be 50 people at the hospital.
- The selected person may not want to participate.

(b) Give **two** advantages of quota sampling.

(2 marks)

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

- Quota sampling is efficient
- Quota sampling ensures better planning
- and uses random techniques.
- and allows for comparison between in-costume visitors and plain-clothed visitors.

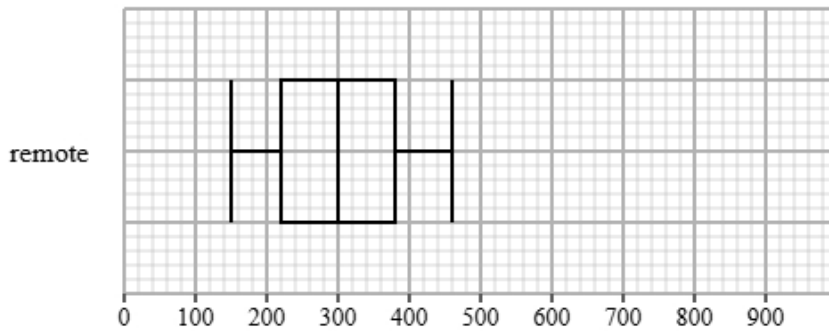
(c) Explain why the quota sample used by Jake is not a random sample.

(1 mark)

Select **one** box.

- The participants volunteered to take part.
- Only the people who go to that entrance are considered.
- Jake should have only used in-costume visitors.
- It is biased.

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

8 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.

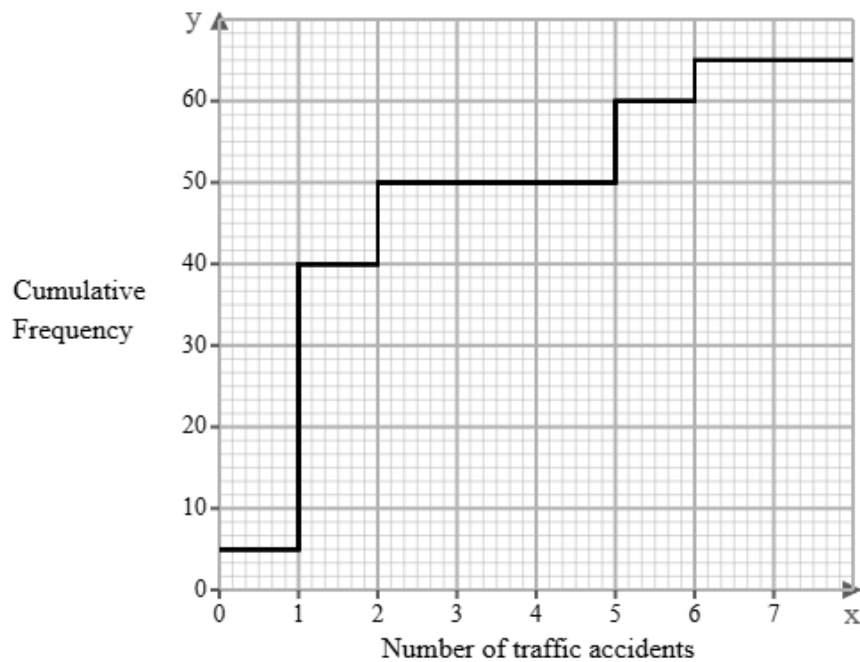
Select **one** box.

- 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.

- 9 The cumulative frequency step polygon shows information about number of traffic accidents reported in a neighbourhood over 65 days.



- (a) Find the mode of the number of traffic accidents.

(1 mark)

The mode is the number that came up the most (the highest frequency).
Look at the cumulative frequency step polygon and see where it 'jumps up' the most.

(b) Find the number of days where there were:

- i) exactly 4 traffic accidents.
- ii) more than 4 traffic accidents.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point.
Remember, the overall frequency is 65.

i) Exactly 4 traffic accidents: _____

ii) More than 4 traffic accidents: _____

(c) In 60 days fewer than x traffic accidents were reported.

Find the value of x

(1 mark)

Draw a line across from 60 on the graph and see where all the 'jumps up' to this line are under.

(d) Rosemary believes the interquartile range of number of traffic accidents reported is 8.

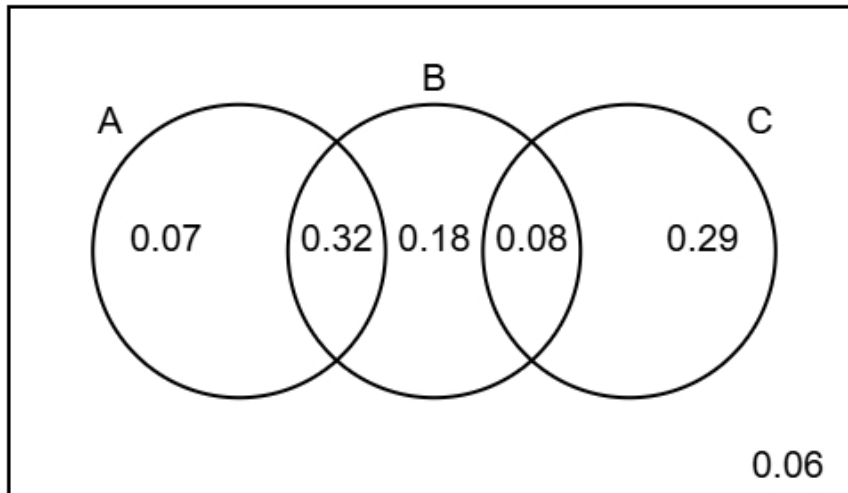
Explain why the interquartile range for this data cannot be 8.

(1 mark)

Select **one** box.

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

10 The Venn diagram illustrates the probabilities associated with events A, B, and C.



(a) Identify the **two** events that are mutually exclusive, giving a reason for your answer. (2 marks)

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

- because they have the lowest total probability.
- because they only overlap once.
- because they do not overlap.
- B and C are mutually exclusive
- A and C are mutually exclusive
- A and B are mutually exclusive

(b) Find $P(B)$

(1 mark)

We are looking for the probabilities inside B.

(c) Find $P(A \text{ or } C)$

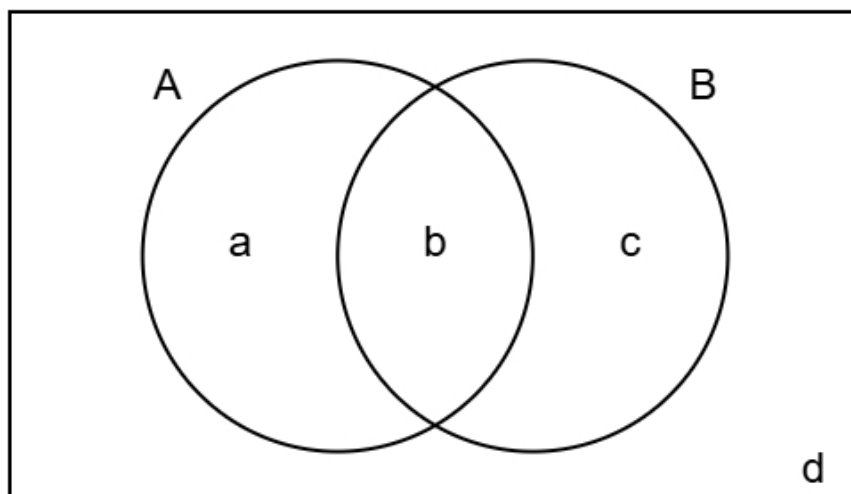
(2 marks)

We are looking for the probabilities inside A or C.

(d) Complete the Venn diagram to show **only** the probabilities for events A and B.

(2 marks)

Combine the probabilities from C into either B or the outside area.



$a =$ _____ $b =$ _____

$c =$ _____ $d =$ _____

Henry is researching the average wait time and customer satisfaction for 10 theme park rides.

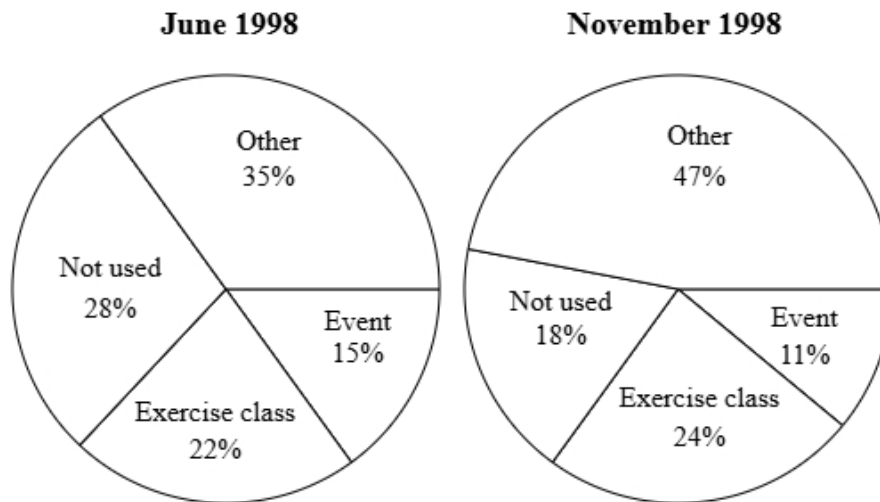
- 11 Discuss whether Henry should have used Pearson's product moment correlation coefficient instead of Spearman's rank correlation coefficient to measure the correlation.

(3 marks)

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

- therefore either would be suitable.
- Henry should use PMCC because
- the PMCC only measures linear correlation
- the PMCC also measures correlation
- Henry should not use PMCC because
- and Spearman's rank correlation is used for ranked data.

12 The pie charts show the uses for a village hall in June 1998 and November 1998.



(a) The number of people using the village hall for exercise classes in June 1998 was 7832.

Find the number of people using the village hall for events in June 1998.

(2 marks)

The number of people using the village hall for exercise classes in June 1998 was 7832 which is shown by 22% on the pie chart.

$$22\% = 7832$$

Find 1%

Then find the number of people using the village hall for events (15%).

(b) In June 1998 the total number of people using the village hall was 36000 (nearest thousand).

In November 1998 the total number of people using the village hall was 43000 (nearest thousand).

Joe wants to use the totals to draw pie charts.

Explain, giving reasons, how Joe can use the totals to draw these pie charts.

(5 marks)

Select **one** box.

- Joe can use a box plot.
- Joe can use comparative pie charts.
- Joe can use dual pie charts.
- Joe can use 3D pie charts.

Select **one** box.

- $\frac{43000^2}{36000^2}$
- $\sqrt{\frac{43000}{36000}}$
- $\left(\frac{43000}{36000}\right)^2$
- $\frac{43000}{36000}$

The radius of the June 1998 pie chart will be _____ (2 d.p.) larger than the November 1998 pie chart.

Select **one** box.

- Because the totals differ across these pie charts, the overall area of each will reflect that difference, while the proportions remain the same.
- With areas inversely related to frequency, the charts become easier to interpret.
- Each sector's size will shift depending on the total amount represented in the chart.
- Compared to other pie chart formats, this method is much faster.

13 Changes in the cost of living in the United Kingdom are measured by the Consumer Price Index (CPI).

The table below shows the index numbers for 2013, using 2005 as the base year, for the items that contribute to the CPI.

It also shows the weightings representing how the Taylor family's spending is distributed among these items.

Item	Index Number (2013)	Taylor Family Weighting (%)
Food, alcoholic beverages & tobacco	113.9	17
Clothing and footwear	110.1	5
Housing and household services	110.5	32
Transport	110.6	13
Recreation, culture and communication	110.5	25
Health, education and other	110.2	8

Using 2005 as base year, the national CPI for 2013 was 113.5.

Compare the total variation in the Jones family's cost of living from 2015 to 2013 with the change in the Consumer Price Index (CPI) over the same period and explain whether your calculations allow you to determine if the Jones family is financially better off or worse off in 2013 compared to 2015.

(5 marks)

Taylor family weighted mean = a (1 decimal place)
 $a < 113.5$

a = _____

*Select **one** box.*

- Cost of living has risen less for the Taylor family.
- Cost of living has risen more for the Taylor family.
- Cost of living has risen in line with the national CPI for the Taylor family.

*Select **one** box.*

- We do not know about the Taylor family's income so cannot determine if they are better off.
- The Taylor family are better off because they have a better cost of living.
- The Taylor family are worse off because they have a better cost of living.

14 A company applies a protective coating to metal parts.

The coating has a target thickness of 0.15 mm.

The company uses quality assurance to monitor the thickness of the coating on each part.

Samples of the parts are taken from the production line at regular intervals and the mean thickness of the coating in each sample is found.

The sample means should be normally distributed with a mean of 0.15 mm and a standard deviation of 0.01 mm.

(a) Find the upper action limit for the sample means for the parts.

(2 marks)

Use the formula for the upper action limit

$$\text{Upper action limit} = \mu + 3\sigma$$

_____ mm

(b) The upper action limit will be set closer to the target thickness of 0.15 mm.

Describe the effect this will have on the frequency of production process stoppages.

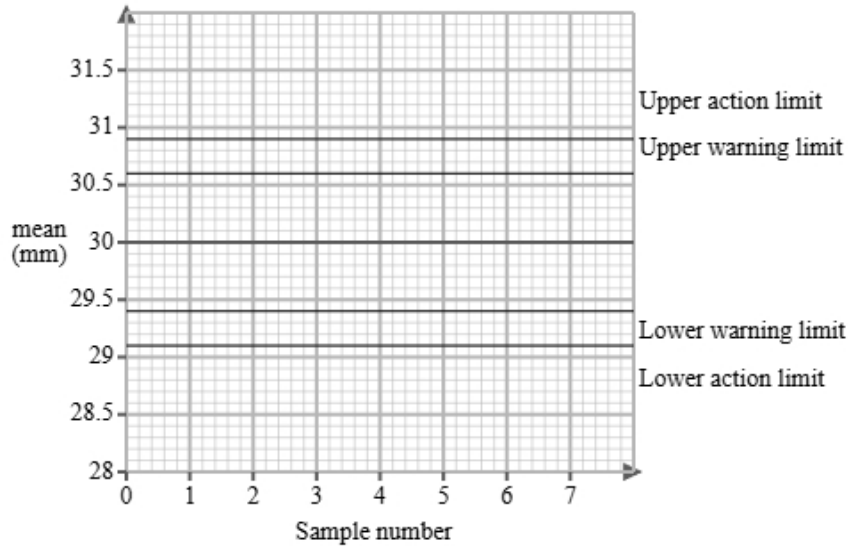
(1 mark)

Select **one** box.

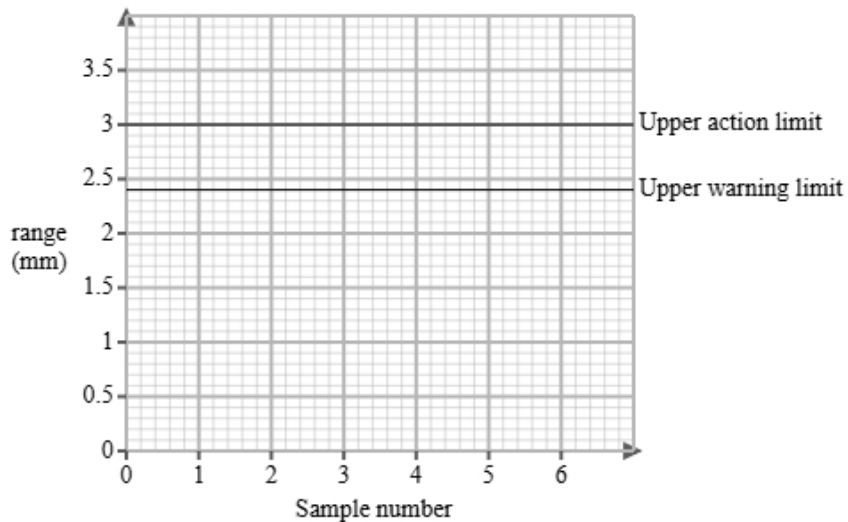
- The production process would be stopped less frequently.
- This is unrelated to the production process so will have no effect.
- The production process would be stopped more frequently.
- The production process will work faster.

- (c) The company also applies a protective coating to plastic parts and uses quality assurance to monitor the thickness of the coating.
 Here are the control charts for the sample means and for the sample ranges of the thicknesses of each plastic part's coating.

Control chart for means



Control chart for ranges



A sample is taken and is found to have a mean of 29.3 mm and a range of 3.1 mm.
 Use the sample mean and range to determine what action, if any, needs to be taken.

(2 marks)

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

- The production process should be stopped as the range is outside the upper action limit
- The production process should continue because the range is within the warning limit
- and the mean is within the upper warning limit.
- although the mean is outside the upper warning limit but not the action limit.

15 A study took place in Canada to find if there was a relationship between hours worked and calories burned of remote workers.

The researchers found the equations of the regression lines for the relationship between hours worked (x hours) and calories burned (y kcal) for junior employees and senior employees the first quarter (Q1) and the second quarter (Q2).

The table below gives the equations of the regression lines.

	Q1	Q2
junior employees	$y = -30x + 2200$	$y = -25x + 2300$
senior employees	$y = -40x + 2400$	$y = -35x + 2500$

(a) Compare the relationships between hours worked and calories burned in junior and senior employees. Include in your comparisons reference to whether it is Q1 or Q2.

(3 marks)

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

- As hours worked increased, the calories burned decreased more rapidly in Q1 compared to Q2.
- As hours worked increased, the calories burned decreased more rapidly in Q2 compared to Q1.
- For all remote workers, an increase in hours worked led to an increase in calories burned.
- The calories burned of junior employees decreased more per hours worked than senior employees.
- For all remote workers, an increase in hours worked led to a decrease in calories burned.
- The calories burned of senior employees decreased more per hours worked than junior employees.

(b) The researchers would like to use a normal distribution as a model for the calories burned of junior employees in Q1.

i) Explain how they could check whether a normal distribution is a suitable model by drawing a histogram.

ii) Explain how they could check whether a normal distribution is a suitable model by calculating the averages and the standard deviation.

(3 marks)

Select **one** box.

If the histogram increases, a normal distribution could be a suitable model.

If the histogram has a skew, a normal distribution could be a suitable model.

If the histogram is symmetrical, a normal distribution could be a suitable model.

If the histogram decreases, a normal distribution could be a suitable model.

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

If the skewness is large (calculated from the mean, median and standard deviation)

and 63% of data is within 2 standard deviations from the mean, a normal distribution could be a suitable model.

If the skewness is 0 (calculated from the mean, median and standard deviation)

and 95% of data is within 2 standard deviations from the mean, a normal distribution could be a suitable model.

16 Each visitor to a website has a 10% chance of making a purchase.

A sample of 4 visitors are randomly selected, and the number of purchases are recorded.

(a) Identify two conditions needed so that a binomial distribution is a suitable model for the number of purchases are recorded.

(2 marks)

Select *two* boxes.

There are only two possible outcomes, purchase or not purchase.

The purchases have a high value.

The probability increases with every purchase.

The chance of a purchase remains constant.

(b) Calculate the probability, as a fraction, that all 3 of the visitors purchased an item.

(2 marks)

Work out p^n where p is the given probability and n is the amount.

(c) Calculate the probability, as a fraction, that at least 2 of the visitors purchased an item.

(3 marks)

You can use Pascal's triangle, or your calculator to find $P(X < 2)$ then take this away from 1.
