statsGCSE

Scaffolded version

Statistics GCSE

Paper 1

Edexcel Higher - 2026

Higher Tier

Variant 1 (same as video) 1ST0/1H

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.

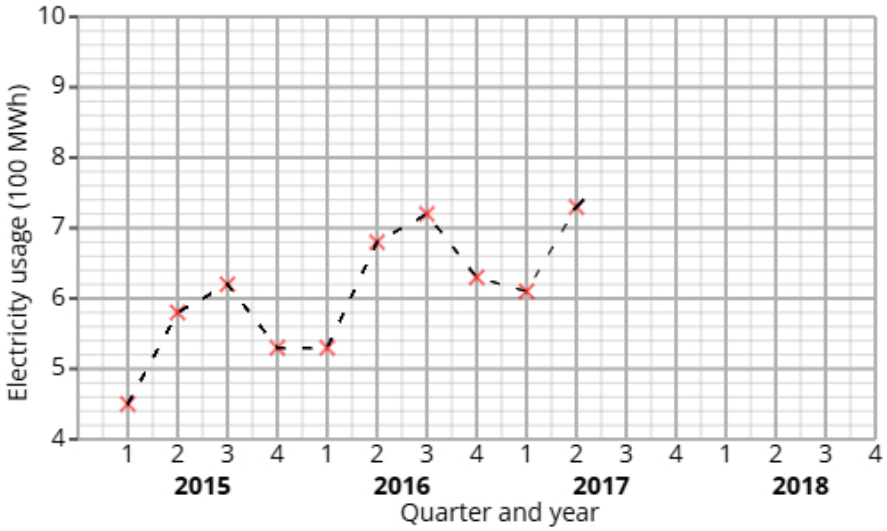
Information

- 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 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) Noah uses the time series graph to estimate that there was 700MWh of electricity usage in Q1 2018

i) Plot the moving averages onto the time series graph and draw a trend line from 2015 to 2017.

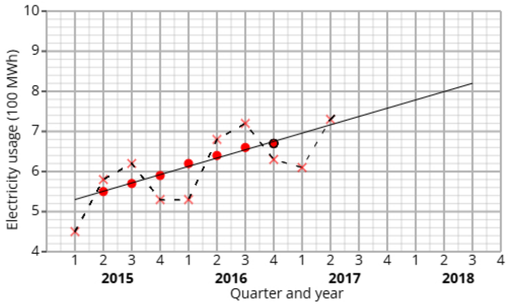
ii) Describe the trend.

iii) By using the average seasonal effect for Q1, show that Noah's estimate is reasonable.

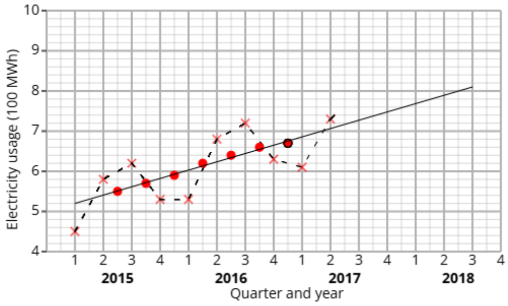
(7 marks)

Select the correct answer.

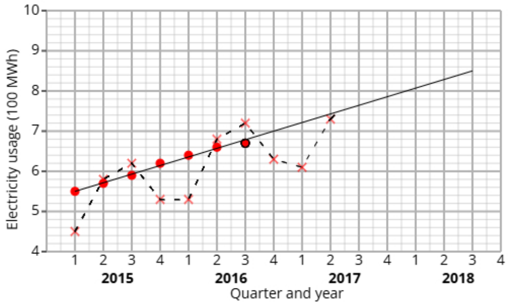
☐



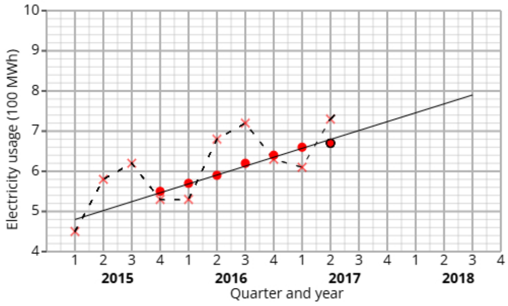
☐



☐



☐



*Select **one** box.*

- ☐ No trend
- ☐ Upward trend
- ☐ Downward trend
- ☐ Flat trend

Select the correct answer.

- ☐ $\frac{770 + 520 + 600 + 680}{4} = 700$
- ☐ $770 - \frac{(520 - 450) + (600 - 530) + (680 - 610)}{3} = 700$
- ☐ $770 - \frac{450 + 530 + 610}{20} = 700$
- ☐ $\frac{770 + 450 + 530 + 610}{3} = 700$

(c) 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.

2 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

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

☐

5-sided spinner

2-sided spinner

	1	**2**	**3**	**4**	**5**
1	2	3	5	6	7
2	3	5	6	7	8

☐

5-sided spinner

2-sided spinner

	1	**2**	**3**	**4**	**5**
1	2	3	4	5	6
2	3	4	5	6	7

☐

5-sided spinner

2-sided spinner

	1	**2**	**3**	**4**	**5**
1	2	3	3	4	5
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

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

3 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.

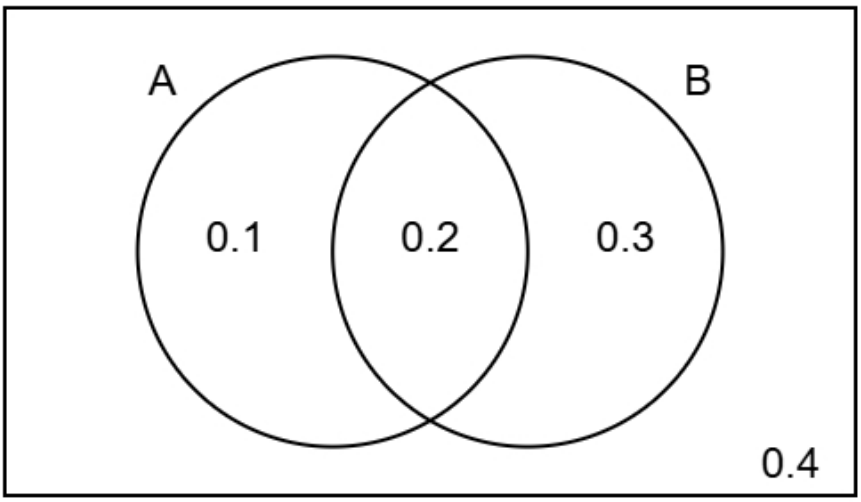
☐ The risk of failing the baking test having taken Workshop X is greater than the risk of failing the baking test having taken Workshop Y.

☐ 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.

4 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 and B)

(1 mark)

P(A and B) is shown in the overlap of the Venn diagram

Leave your answer as a decimal.

(c) Find $P(B \mid A)$

(2 marks)

Use the formula to find $P(B \mid A)$

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

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

$P(P) = 0.6$

$P(Q) = 0.7$

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

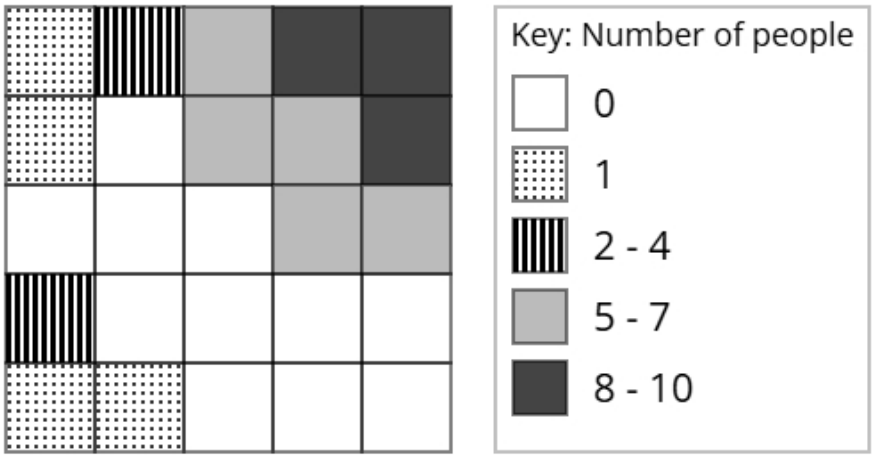
(2 marks)

Use the formula for independent events

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

Leave your answer as a decimal.

5 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.

6 A scientist is conducting an experiment to investigate the effect of a dietary supplement on alertness.

She plans to use a matched pairs design.

Nia is one of the participants in the study.

As part of the experiment, she completes four alertness tests.

Each test has a different weighting.

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

Test	Weighting	Score
A	3	21
B	2	17
C	4	28
D	1	12

(a) Explain the concept of matched pairs in an experimental design.

(2 marks)

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

☐ so that the experiment is less intimating.

☐ Matched pairs is a type of experimental design where participants are paired based on similar characteristics

☐ Matched pairs is a type of experimental design where participants work with a partner in the experiment

☐ so that the experimenter can control for other factors.

(b) Calculate the weighted mean score for Nia's four tests.

(3 marks)

Use the formula

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

7 Samira is trying to estimate the fish population in a lake.

She first captures and tags 60 fish, then releases them.

One week later, she catches a second sample of 24 fish.

Using the Petersen capture-recapture method, she estimates the total number of fish to be 360.

(a) How many of the 360 fish in Samira's second sample were tagged?

(2 marks)

Use the capture recapture formula (you will need to rearrange it)

$$N = \frac{M \times n}{m}$$

where:

N = estimated total population

M = number of individuals originally marked (first capture)

n = total number of individuals captured in the second sample

m = number of marked individuals recaptured in the second sample

(b) Discuss how reliable Samira's estimate is by considering the assumptions required for using the Petersen capture-recapture method.

(3 marks)

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

- ☐ She has also assumed that there were no fish that died, which is unlikely.
- ☐ These assumptions mean that the estimate may not be reliable, because the assumptions are unlikely to hold.
- ☐ Samira has assumed that there are only fish in the lake.
- ☐ This means that the estimate will be reliable.
- ☐ Samira has assumed that she selected a random sample, which is unlikely.
- ☐ She has also collected a very large sample.

8 A hospital administrator wants to find out if nurses have used their phones for non-work-related social media during shifts in the last 3 months.

Oliver suggests using the random response technique to ask the nurses about using their phones for social media during their shifts.

(a) Explain why Oliver has suggested using the random response technique for this situation.

(1 mark)

*Select **one** box.*

☐ Oliver suggested the random response technique to encourage honest answers to a sensitive question.

☐ Oliver used the random response technique to make sure everyone gives the same answer.

☐ Oliver suggested the random response technique so he could skip analysing individual responses.

☐ Oliver chose the random response technique because it guarantees the results will be normally distributed.

(b) The hospital administrator uses the random response question below:

> Think of a month of the year.
>
> If the month is after June, do not read the question and tick box A.
>
> If the month is June or before, answer the question truthfully.
>
> Have you used your phone for non-work-related social media during a shift?
>
> If yes, tick box A. If no, tick box B.
>
> A ☐ B ☐

State if this question is appropriate and give reasons for your answer.

(2 marks)

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

☐ people are randomly picking a month

☐ so no one will know if they were forced to answer A.

☐ the month people will pick is not random

☐ This question is not appropriate because

☐ This question is appropriate because

☐ and there is no time frame in the question.

(c) The final questionnaire will be distributed to a sample of nurses.

The nurses are made up from three fields adult, mental health, and children's nursing.

They work either full-time or part-time.

The table shows how many nurses there are in each category

		Field		
		Adult	Mental Health	Children
Employment status	Full-time	69	77	80
	Part-time	76	41	25

The hospital administrator plans to take a stratified sample based on field and employment status and requires a minimum of 10 individuals from each stratum.

If the calculated sample size for a particular stratum is a decimal, he will round it to the nearest integer.

Determine the smallest total sample size that ensures at least 10 people are selected from each stratum.

(2 marks)

Find the smallest group

Decimals are rounded up, so find the smallest sample for this group before rounding it to 10

Use the stratified sampling formula to find n (the total number required in the sample)

9 Alex is investigating the profits made by two different shops, GreenMart and FreshFields.

Alex has obtained the annual percentage profits made by GreenMart for the years 2012 to 2016 and the annual percentage profits made by FreshFields for the years 2013 to 2016.

The table below gives this information.

Year	Percentage profit (%)	
	GreenMart	FreshFields
2012	2.2	
2013	2.5	1
2014	2.9	1.7
2015	3.3	1.4
2016	3.7	5.9

Alex concludes that the average annual percentage profit made by FreshFields over the 4 years is greater than the average annual percentage profit made by GreenMart over the 5 years.

By using appropriate geometric means, assess Alex's conclusion.

You must show your working.

(5 marks)

$\text{geometric mean} = \sqrt[n]{\text{value}_1 \times \text{value}_2 \times \ldots \times \text{value}_n}$

You will need to convert all the percentage increases into multipliers

*Select **one** box.*

☐ Alex's conclusion is correct.

☐ Alex's conclusion is not correct.

10 Tomi has collected data about the heights, in cm, of students in a school.

The table gives some of the percentiles of Tomi's data.

Percentile	Height (cm)
97.5th	182.5
80th	173.8
60th	169.4
40th	165.6
20th	161.2
5th	155.2
2.5th	152.5

(a) Find the 2.5th to 97.5th interpercentile range.

(1 mark)

The 2.5th to 97.5th interpercentile range is the distance between those percentiles

_____ cm

(b) One of the students from the sample is selected at random.

Find the probability that their height is between 155.2 cm and 169.4 cm.

(1 mark)

Find the percentiles for 155.2 cm and 169.4 cm

The probability will be the distance between the percentiles

_____ %

(c) Give a reason why it is appropriate for Tomi to use the mean and the standard deviation to summarise this data

(1 mark)

*Select **one** box.*

- ☐ The distribution is positive.
- ☐ The distribution is symmetric.
- ☐ The data is numeric.
- ☐ The data is likely to have outliers.

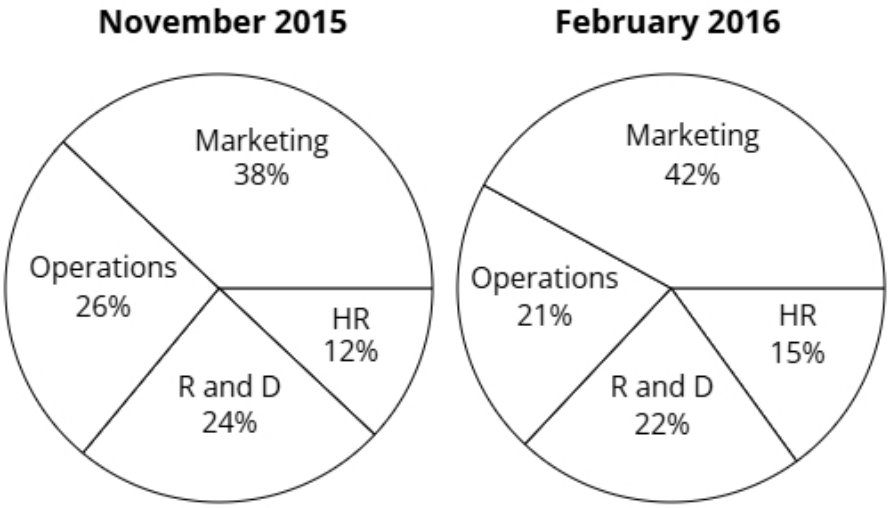
(d) Tomi claims that the heights of the students can be modelled using a normal distribution with mean 167.5 cm and standard deviation 7.5 cm.
Use the data to assess the validity of Tomi's claim.

(4 marks)

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

- ☐ 97.5% – 2.5% = 95% so the claim is valid.
- ☐ $167.5 - 3 \times 7.5 = 145$
- ☐ 167.5 is not shown in the table, so is incorrect.
- ☐ $167.5 - 2 \times 7.5 = 152.5$
- ☐ $167.5 + 3 \times 7.5 = 190$
- ☐ 80% – 20% = 60% so the claim is not valid.
- ☐ The distribution is symmetric so $\frac{169.4 + 165.6}{2} = 167.5$
- ☐ $167.5 + 2 \times 7.5 = 182.5$

11 The pie charts show the budget allocations for a business in November 2015 and February 2016.



In November 2015 the total budget was £155000 (nearest thousand).

In February 2016 the total budget was £185000 (nearest thousand).

Gemma wants to use the totals to draw pie charts.

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

(5 marks)

*Select **one** box.*

☐ Gemma can use comparative pie charts.

☐ Gemma can show the totals instead of the percentages.

☐ Gemma can use dual pie charts.

☐ Gemma could use a bar chart.

*Select **one** box.*

- ☐ $\left(\frac{185000}{155000}\right)^2$
- ☐ $\frac{185000^2}{155000^2}$
- ☐ $\sqrt{\frac{185000}{155000}}$
- ☐ $\frac{185000}{155000}$

The radius of the November 2015 pie chart will be _____ (2 d.p.) larger than the February 2016 pie chart.

*Select **one** box.*

- ☐ This method is much quicker to do than other types of pie chart.
- ☐ The totals for these pie charts are different, so the areas of each pie chart will represent the totals whilst keeping the proportions the same.
- ☐ The areas will now be inversely proportional to the frequency, making the pie charts much easier to understand.
- ☐ The proportions of each sector will change depending on the total for each pie chart.

12 A company bottles soft drinks.

The bottles have a target volume of 500 ml.

The company uses quality assurance to monitor the volume of soft drink in each bottle.

Samples of the bottles are taken from the production line at regular intervals and the mean volume of soft drink in the bottles in each sample is found.

The sample means should be normally distributed with a mean of 500 ml and a standard deviation of 3 ml.

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

(2 marks)

Use the formula for the upper action limit

Upper action limit $= \mu + 3\sigma$

_____ ml

(b) The upper action limit will be set closer to the target volume of 500 ml.

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

(1 mark)

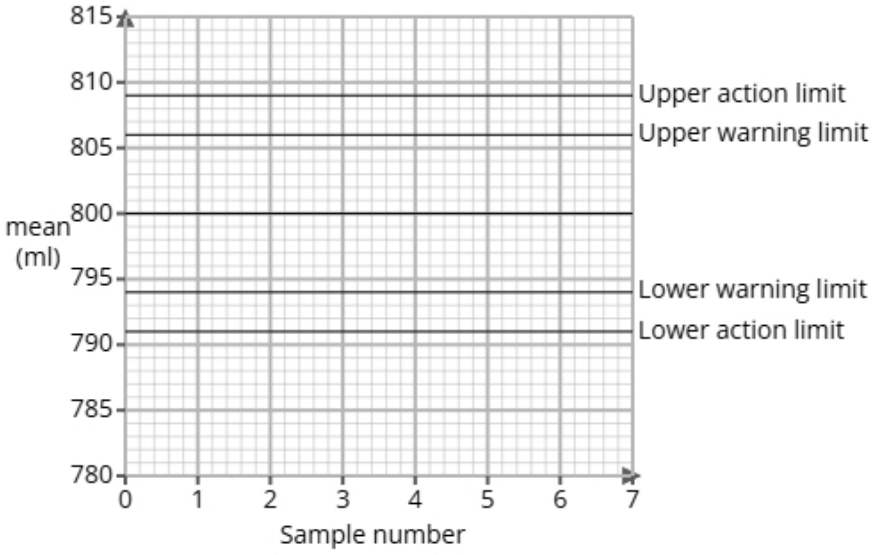
*Select **one** box.*

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

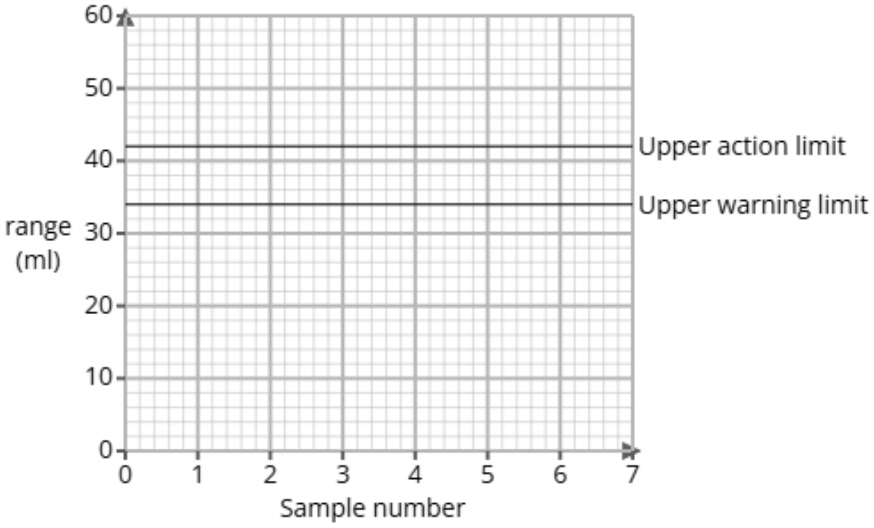
(c) The company also bottles tomato ketchup and uses quality assurance to monitor the volume of tomato ketchup in those bottles.

Here are the control charts for the sample means and for the sample ranges of the volumes of tomato ketchup in the bottles.

Control chart for means



Control chart for ranges



A sample is taken and is found to have a mean of 807 ml and a range of 43 ml.

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

13 A study took place in Sweden to find if there was a relationship between daily exercise and resting heart rate of middle-aged men.

The researchers found the equations of the regression lines for the relationship between daily exercise (x minutes) and resting heart rate (y bpm) for gym members and non-gym members on weekdays and weekends.

The table below gives the equations of the regression lines.

	weekdays	weekends
gym members	$y = -0.3x + 72$	$y = -0.25x + 75$
non-gym members	$y = -0.1x + 78$	$y = -0.05x + 80$

(a) Compare the relationships between daily exercise and resting heart rate in middle-aged men who are gym and non-gym members. Include in your comparisons reference to whether it is a weekday or weekend.

(3 marks)

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

- ☐ For all middle-aged men, an increase in daily exercise led to a increase in resting heart rate.
- ☐ The resting heart rate of gym members decreased more per minutes of daily exercise than non-gym members.
- ☐ As daily exercise increased, the resting heart rate decreased more rapidly on weekends compared to weekdays.
- ☐ The resting heart rate of non-gym members decreased more per minutes of daily exercise than gym members.
- ☐ As daily exercise increased, the resting heart rate decreased more rapidly on weekdays compared to weekends.
- ☐ For all middle-aged men, an increase in daily exercise led to a decrease in resting heart rate.

(b) The researchers would like to use a normal distribution as a model for the resting heart rate of gym members on weekdays.

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 is positively skewed, a normal distribution could be a suitable model.

☐ If the histogram shows an increase, a normal distribution could be a suitable model.

☐ If the histogram is a bell shape, a normal distribution could be a suitable model.

☐ If the histogram shows a decrease, a normal distribution could be a suitable model.

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

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

☐ and 68% of data is within 2 standard deviations from the mean, a normal distribution could be a suitable model.

☐ If the mean, median and mode are equal

☐ If the mean, median and mode are larger than the standard deviation

14 20% of the emails a user receives are spam emails.

Out of 3 incoming emails, the number of spam emails is recorded.

(a) Identify two conditions needed so that a binomial distribution is a suitable model for the number of spam emails in the incoming emails.

(2 marks)

*Select **two** boxes.*

- ☐ Each of the emails are independent.
- ☐ There are only two possible outcomes, spam or not spam.
- ☐ Each email increased the likelyhood of a spam email.
- ☐ The emails arrive at the same time.

(b) Calculate the probability, as a fraction, that all 3 of the incoming emails are spam.

(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 incoming emails are spam.

(3 marks)

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