Statistics GCSE

Paper 2

2025

Edexcel Higher

Variant 5 — 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 At a career fair, 90% of attendees are recent graduates and 10% are professionals. Ella and Nathan plan to conduct a feedback survey.

Ella decides to use simple random sampling to select 100 participants. She uses the list of registered attendees as a sampling frame, assigning a number to each participant. She then generates 100 random numbers and selects her sample accordingly.

Nathan decides to use quota sampling to collect a sample of 100 participants. He plans to stand near one of the entrances until 90 recent graduates and 10 professionals have been interviewed.

(a) Give two reasons why Ella's method may **not** produce a sample of 100 participants.

(2 marks)

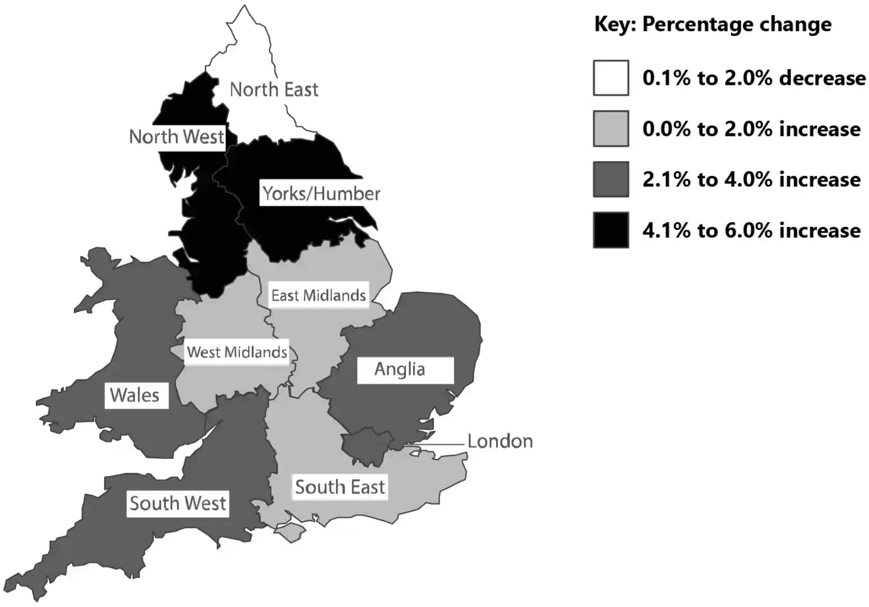
(b) Give **two** advantages of quota sampling.

(2 marks)

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

(1 mark)

2 The map below shows the percentage change in domestic tourist visits across different regions of England and Wales between 2014 and 2015.



(a) Write down the percentage change in domestic tourist visits between 2014 and 2015 in North West.

(1 mark)

____

____

____

____

(b) There are 10 regions shown.
Find the number of regions that the domestic tourist visits **decreased**.

(1 mark)

____

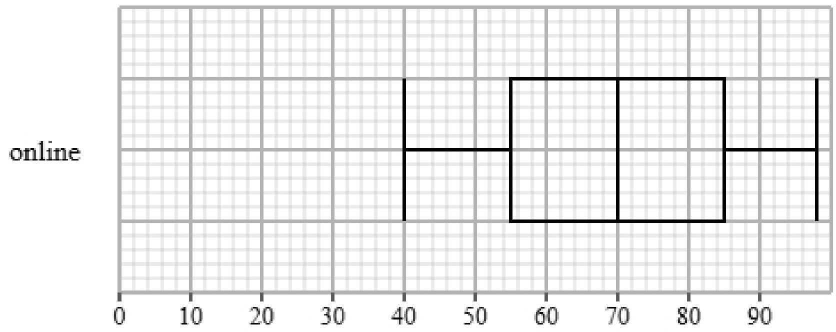
(c) Hadley states that domestic tourism in England and Wales increased overall between 2014 and 2015.
Explain why this may not be the case.

(1 mark)

(d) State the statistical name for the type of map shown.

(1 mark)

3 Isla collected the satisfaction scores for online and in-store customers.
Both types of shopping experiences were rated by the same number of customers.
The box plot presents data on the satisfaction scores for the online customers.

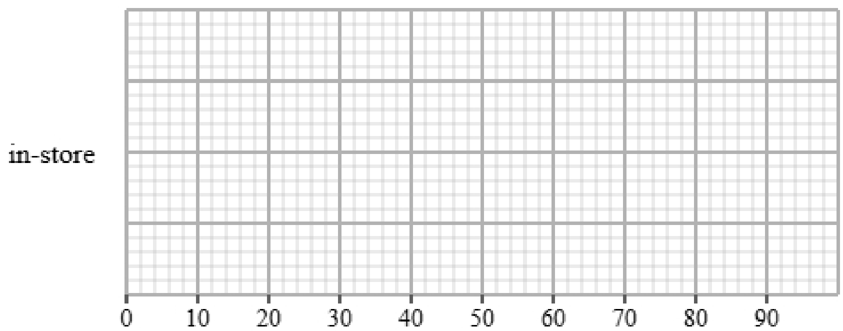


The table gives information about the satisfaction scores for the in-store customers.

Least tall	Lower quartile	Median	Upper quartile	Most tall
40	60	80	85	90

(a) Draw a box plot for the satisfaction scores for the in-store customers.

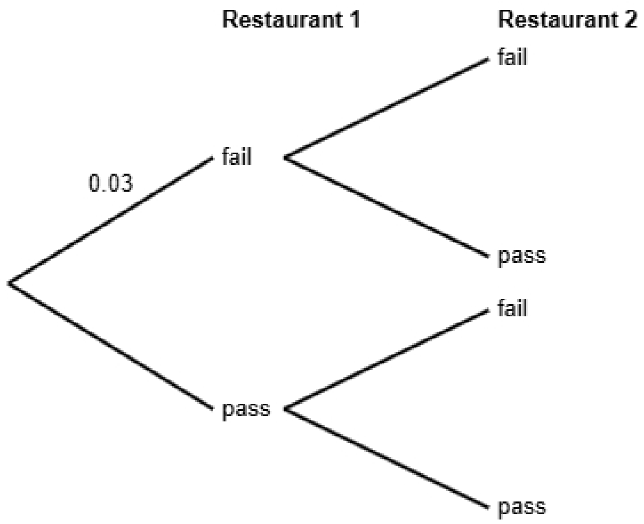
(2 marks)



(b) Compare the two distributions of satisfaction scores.
Give three comparisons and interpret one of these comparisons.

(4 marks)

4 Health inspections show that 3% of restaurants in a city fail basic hygiene standards.
The rest pass the required health standards.
Emma is a food safety officer and will be inspecting two different restaurants.
She does not know if each restaurant will fail or pass.



(a) Complete the probability tree diagram.

(2 marks)

(b) Find the probability that both of Emma's restaurants have an outcome of pass.

(2 marks)

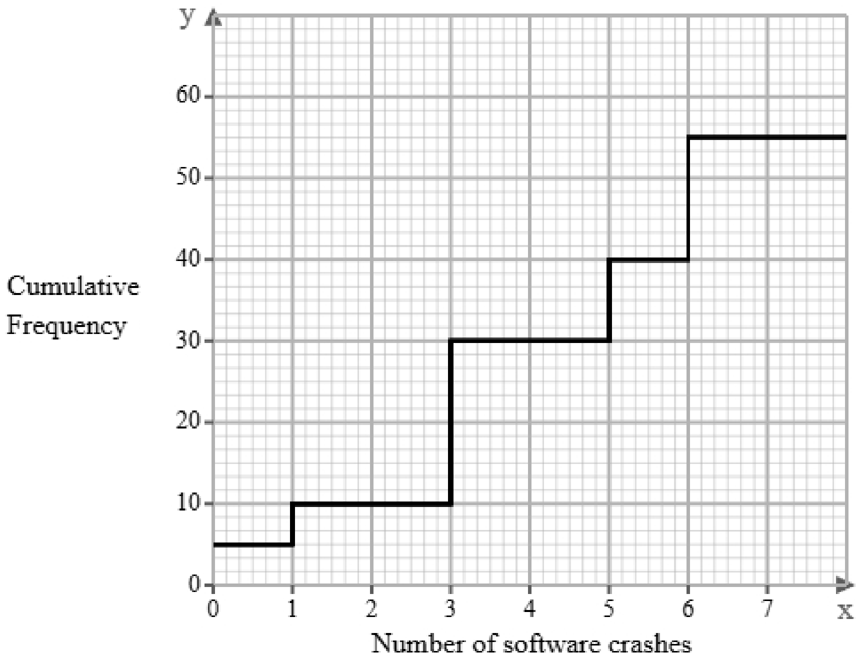
____

(c) Emma states that the probability that exactly one restaurant outcome is fail is less than 6%

Find out whether or not Emma is correct.

(3 marks)

5 The cumulative frequency step polygon shows information about number of software crashes experienced by a company's servers over 55 days.



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

(1 mark)

____

____

____

____

(b) Find the mode of the number of software crashes.

(1 mark)

____

(c) Find the number of days where there were:
i) exactly 4 software crashes.
ii) more than 4 software crashes.

(3 marks)

i) Exactly 4 software crashes: __________

ii) More than 4 software crashes: __________

(d) In 40 days fewer than x software crashes were experienced.
Find the value of x

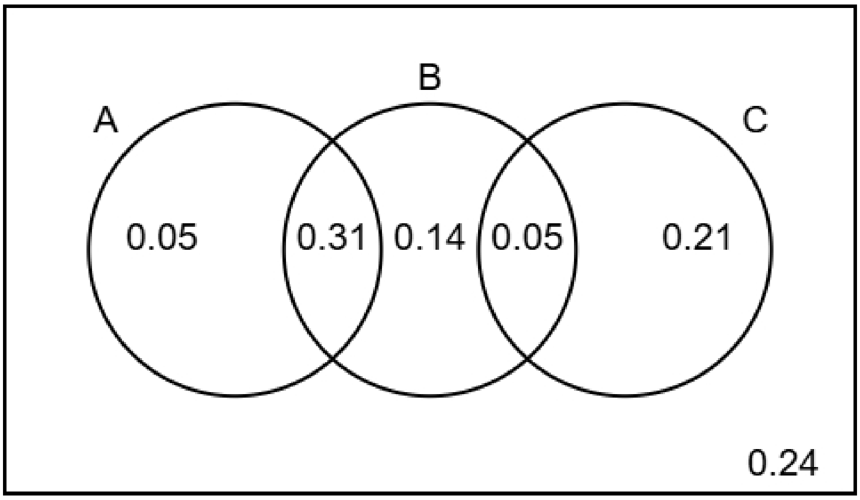
(1 mark)

____

(e) Delia believes the interquartile range of number of software crashes is 8.
Explain why the interquartile range for this data cannot be 8.

(1 mark)

6 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)

(b) Find P(B)

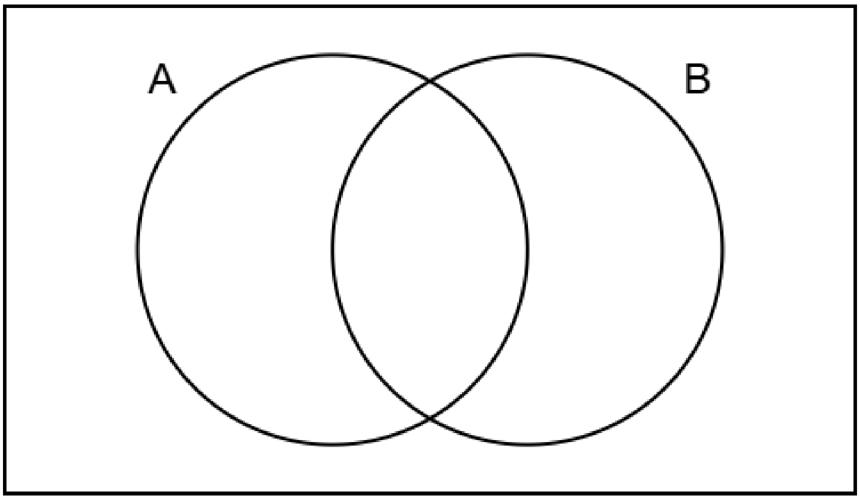
(1 mark)

(c) Find P(A or C)

(2 marks)

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

(2 marks)



7 Sebastian is researching average daily number of cups sold and the coffee quality rating for 10 coffee blends at a coffee shop.

(a) Suggest a diagram that Sebastian could draw to determine if there is a relationship between average daily number of cups sold and the coffee quality rating for the 10 coffee blends at a coffee shop.

(1 mark)

(b) i) Calculate Spearman's rank correlation coefficient from the table.
ii) Interpret your answer in the context of Sebastian's research. You will need to refer to the effects of any outliers.

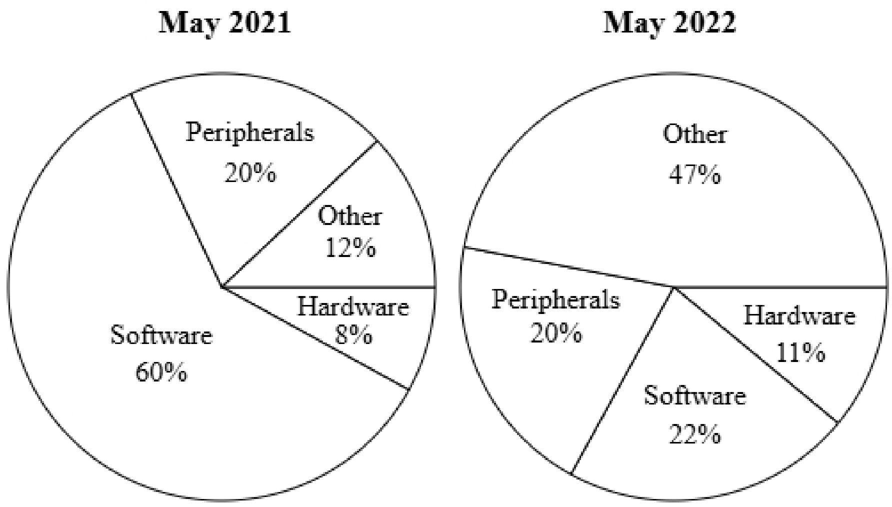
(5 marks)

Coffee Blend	Average Daily Number of Cups Sold	Cups Sold Rank	Coffee Quality Rank	d	d^2
Caramel Delight	92	1	2	-1	1
Caramel Macchiato	21	8	7	1	1
Choco Hazelnut	72	3	3	0	0
Coconut Crunch	28	7	8	-1	1
Dark Roast Supreme	35	6	6		
Hazelnut Dream	89	2	5		
Mocha Fusion	61	5	4		
Mocha Java	2	10	9		
Vanilla Bean Bliss	63	4	6		
Espresso Royale	9	9	10		

(c) Discuss whether Sebastian should have used Pearson's product moment correlation coefficient instead of Spearman's rank correlation coefficient to measure the correlation.

(3 marks)

8 The pie charts show the types of sales at a computer store in May 2021 and May 2022.



(a) Explain why the pie charts do not show an increase in hardware sales in May 2022 compared to May 2021.

(1 mark)

____

____

____

____

(b) The sales for software in May 2021 was £59160.
Find the sales for hardware in May 2021.

(2 marks)

£__________

(c) In May 2021 the total sales was £99000 (nearest thousand).
In May 2022 the total sales was £132000 (nearest thousand).

Maria wants to use the totals to draw pie charts.

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

(5 marks)

9 A company manufactures LED light bulbs.
The bulbs have a target brightness of 800 lumens.
The company uses quality assurance to monitor the brightness of each bulb.
Samples of the bulbs are taken from the production line at regular intervals and the mean brightness of the bulbs in each sample is found.
The sample means should be normally distributed with a mean of 800 lumens and a standard deviation of 10 lumens.

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

(2 marks)

__________ lumens

(b) The upper action limit will be set closer to the target brightness of 800 lumens.
Describe the effect this will have on the frequency of production process stoppages.

(1 mark)

____

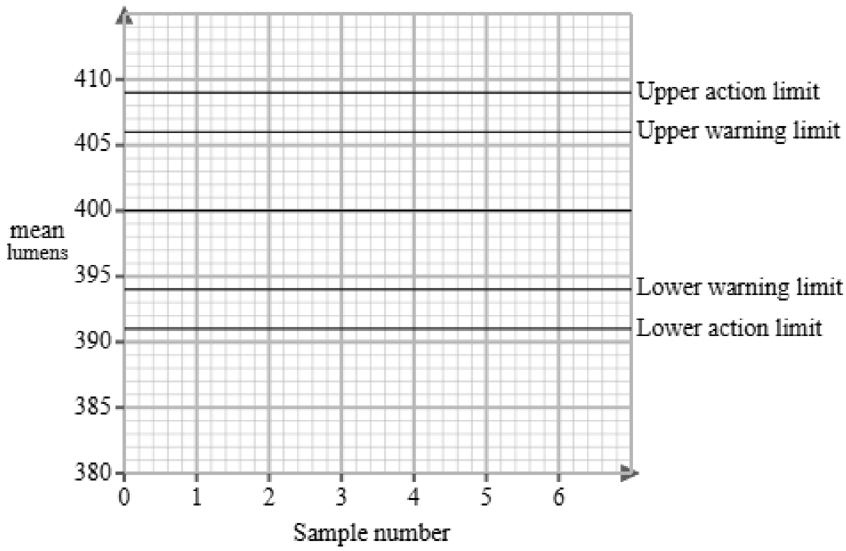
____

____

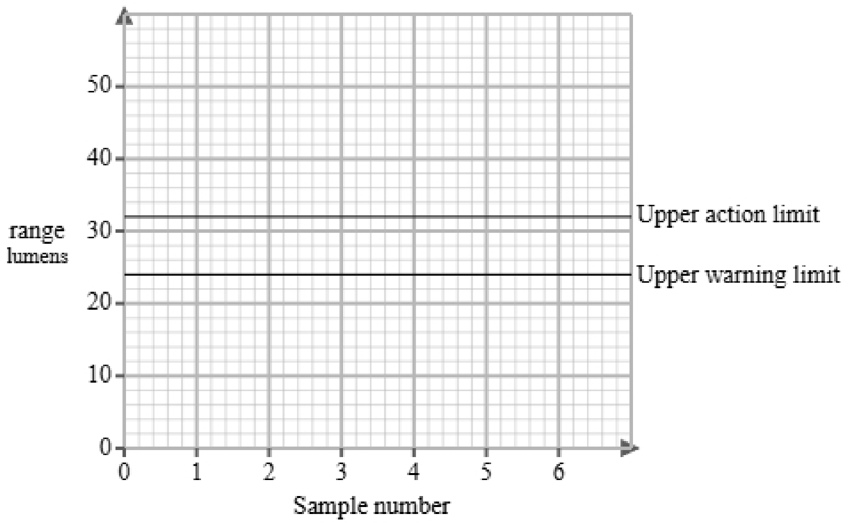
____

(c) The company also manufactures LED light strips and uses quality assurance to monitor the brightness of each strip.
Here are the control charts for the sample means and for the sample ranges of the brightness of each of the strips.

Control chart for means



Control chart for ranges



A sample is taken and is found to have a mean of 408 lumens and a range of 32 lumens.
Use the sample mean and range to determine what action, if any, needs to be taken.

(2 marks)

10 The running speeds of a group of athletes have a mean of 14.5 km/h and a standard deviation of 1.8 km/h.

(a) Olivia is athlete with a standardised score of 0.
Find Olivia's running speed.

(1 mark)

__________ km/h

(b) Aisha and Mia are both athletes in the group.
Aisha's standardised score for running speed is 1.3 km/h.
Mia's standardised score for running speed is -0.7 km/h.
Aisha was faster than Mia.
How much faster is Aisha?

(3 marks)

__________ km/h

(c) Emily takes a sample of 5 athletes from the group.
She wants to calculate the standardised score for the sample mean of their ages.

(i) Discuss the appropriateness of using 14.5 km/h as the mean in the calculation of the standardised score,
(ii) Discuss the appropriateness of using 1.8 km/h as the standard deviation in the calculation of the standardised score.

(4 marks)

11 A study took place in Australia to find if there was a relationship between age and number of books read of public library members.
The researchers found the equations of the regression lines for the relationship between age (x years) and number of books read (y books) for urban members and rural members who prefer fiction and non-fiction.

The table below gives the equations of the regression lines.

	prefer fiction	prefer non-fiction
urban members	$y = -0.2x + 25$	$y = -0.1x + 20$
rural members	$y = -0.3x + 30$	$y = -0.15x + 22$

(a) Interpret in context the figure -0.15 in the regression equation for rural members who prefer non-fiction.

(1 mark)

(b) Compare the relationships between age and number of books read in urban and rural members. Include in your comparisons reference to whether they prefer fiction or non-fiction.

(3 marks)

(c) The researchers would like to use a normal distribution as a model for the number of books read of urban members who prefer fiction.

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)

12 A spinner has 8 equal sections, numbered 1 to 8.
It is spun 4 times.
The number of times it lands on a 8 is recorded.

(a) Identify two conditions needed so that a binomial distribution is a suitable model for the number of times an 8 was recorded.

(2 marks)

(b) Calculate the probability, as a fraction, that all 4 of the spinners land on an 8.

(2 marks)

(c) Calculate the probability, as a fraction, that at least 2 of the spinners land on an 8.

(3 marks)