Statistics GCSE

Paper 2

2025

Edexcel Higher

Variant 5 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

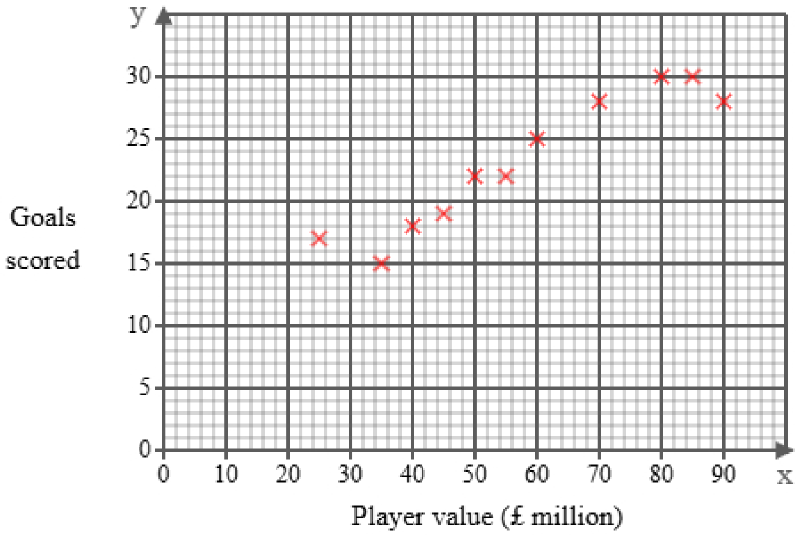
- This is a practise paper to aid your revision for your exams.
- This site, and all that work on it, have no affiliation or relationship with any exam board.
- This site is not endorsed by any company or charity, unless we state otherwise.

Copyright

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 Alice collected data on 11 football strikers, recording the value (in millions of pounds) and the goals scored last season of each player. She represented her findings in the scatter diagram below.



(a) One of the 11 strikers scored 15 goals last season.
Write down the value of the striker.

(1 mark)

£ ____________________

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

(1 mark)

(c) Describe and interpret the type of correlation shown by the scatter diagram.

(3 marks)

(d) A new strike has just been signed for value of £150 million.
Alice is planning on using the line of best fit on the scatter diagram to predict the goals the strike will score this season.

Explain whether or not it is appropriate to use the line of best fit for this prediction.

(2 marks)

A theme park is considering building a new roller coaster.

Mia is carrying out a survey to see what all visitors think about the new attraction.

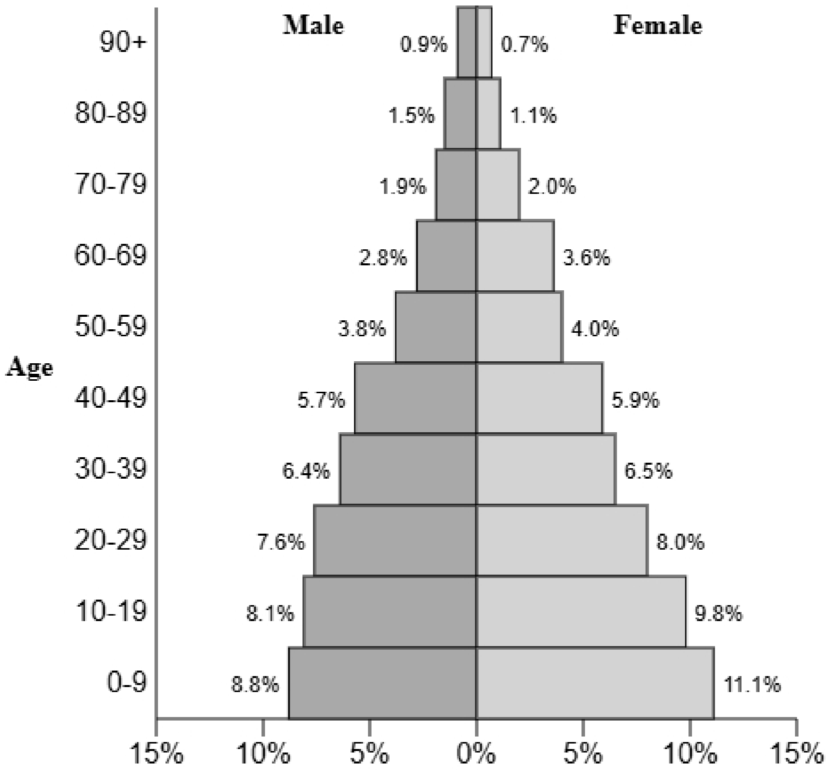
Mia thinks that she should take a sample rather than a census.

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

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

(1 mark)

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



(a) Find the age group that has 15.6% of the population.

(1 mark)

____

____

____

____

(b) Compare the percentage of the population aged 40-69 between males and females.

(1 mark)

____

____

____

____

(c) Give a reason why the sum of all the percentages is 100.2% and not 100%.

(1 mark)

A scientist is conducting an experiment to investigate the effect of background music on learning.
She plans to use a matched pairs design.
Ben is one of the participants in the study.
As part of the experiment, he completes four learning assessments.
Each test has a different weighting.
The table below shows the weightings and Ben's scores for each test.

Test	Weighting	Score
A	2	16
B	4	29
C	1	10
D	3	21

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

(3 marks)

5 Mei works for a marine conservation group. She has been tasked with investigating plastic waste collected by divers.
Below is a section of the spreadsheet she used to record her findings.

Waste (kg)	Percentage of dive logs
$0 < p \leqslant 2$	7
$2 < p \leqslant 4$	seven
$4 < p \leqslant 6$	6
$6 < p \leqslant 8$	118
$8 < p \leqslant 10$	47
$10 < p \leqslant 12$	15
Total	100

Mei cleans the data to create the table below.

Waste (kg)	Percentage of dive logs
$0 < p \leqslant 2$	7
$2 < p \leqslant 4$	7
$4 < p \leqslant 6$	6
$6 < p \leqslant 8$	18
$8 < p \leqslant 10$	47
$10 < p \leqslant 12$	15
Total	100

(a) Give a reason Mei cleaned the data.

(1 mark)

(b) Mei realised that the value of 118 in the original table was incorrect.
Explain how Mei knew this.

(1 mark)

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

(3 marks)

__________ kg

6 The table shows the average monthly platinum price, in pounds, over a six-month period in 2018.
It also provides selected chain base index numbers, rounded to one decimal place, for the same data.

	Jan	Feb	Mar	Apr	May	Jun
platinum price	850	845	860	870	855	865
Chain base index number		99.4	101.8	101.2	98.3	

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

(2 marks)

____

(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)

7 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)

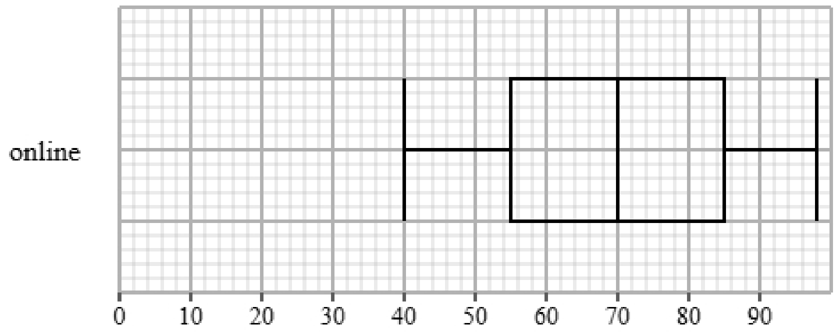
(b) Give **two** advantages of quota sampling.

(2 marks)

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

(1 mark)

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.



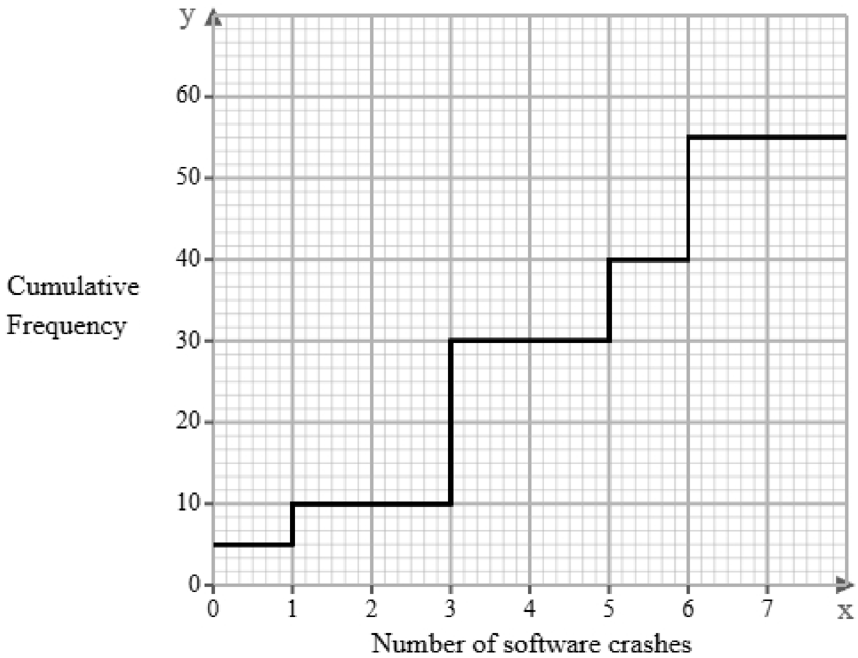
The table gives information about the satisfaction scores for the in-store customers.

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

8 Compare the two distributions of satisfaction scores.
Give three comparisons and interpret one of these comparisons.

(4 marks)

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



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

(1 mark)

____

(b) 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: __________

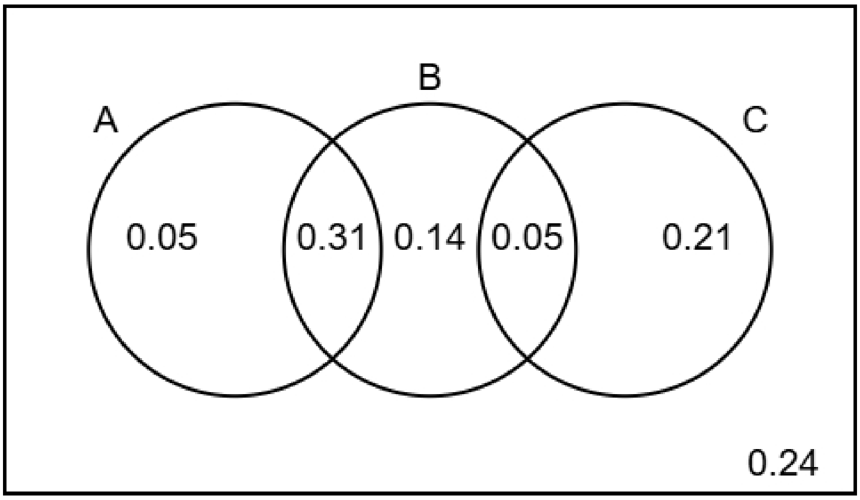
(c) In 40 days fewer than x software crashes were experienced.
Find the value of x

(1 mark)

(d) 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)

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)

(b) Find P(B)

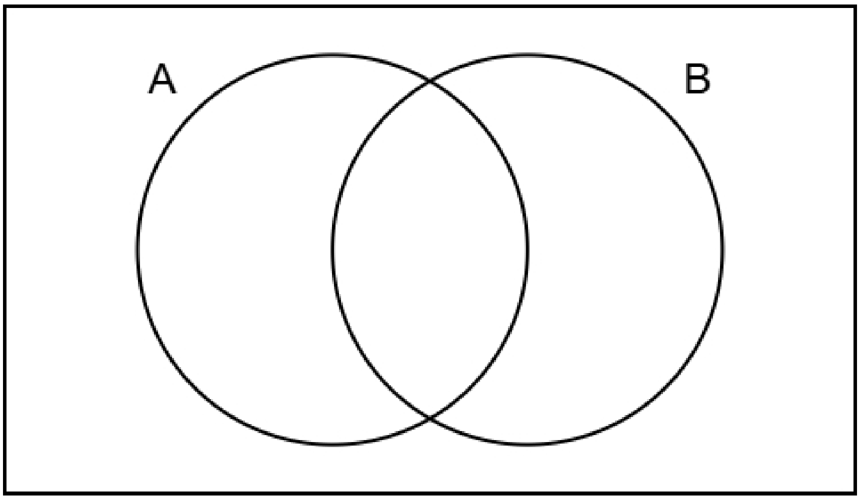
(1 mark)

(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)

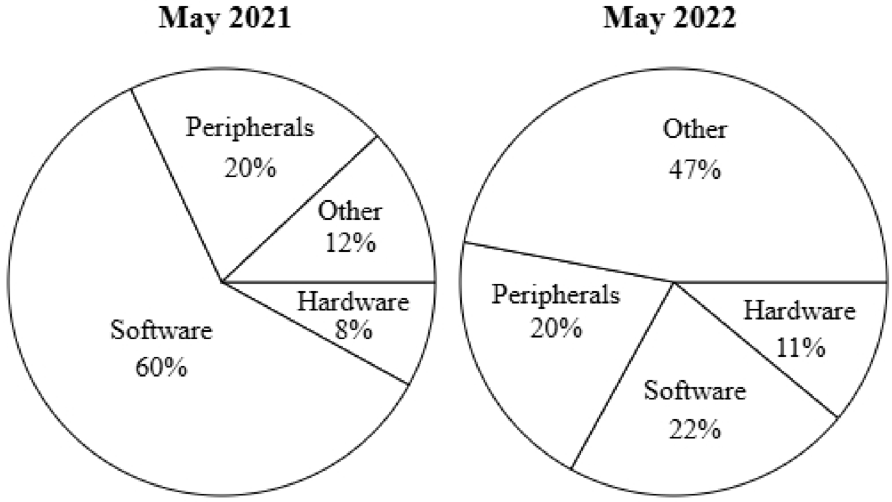


Sebastian is researching average daily number of cups sold and the coffee quality rating for 10 coffee blends at a coffee shop.

11 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)

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



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

(2 marks)

£__________

(b) 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)

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 2014, using 2005 as the base year, for the items that contribute to the CPI.
It also shows the weightings representing how the Jones family's spending is distributed among these items.

Item	**Index Number (2014)**	**Jones Family Weighting (%)**
Food, alcoholic beverages & tobacco	114.6	18
Clothing and footwear	110.9	8
Housing and household services	111.2	27
Transport	111.1	14
Recreation, culture and communication	111.0	24
Health, education and other	110.6	9

Using 2005 as base year, the national CPI for 2014 was 114.6.
Compare the total variation in the Jones family’s cost of living from 2015 to 2014 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 2014 compared to 2015.

(5 marks)

14 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)

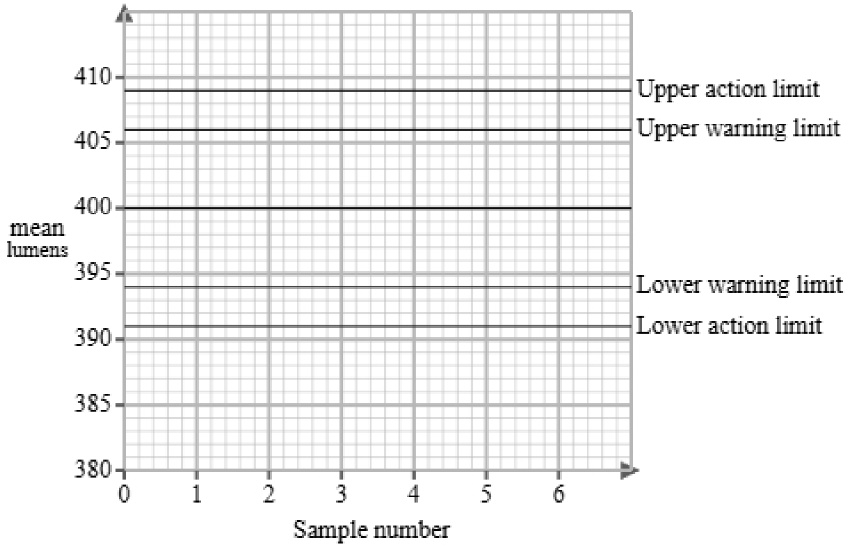
__________ lumens

(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.

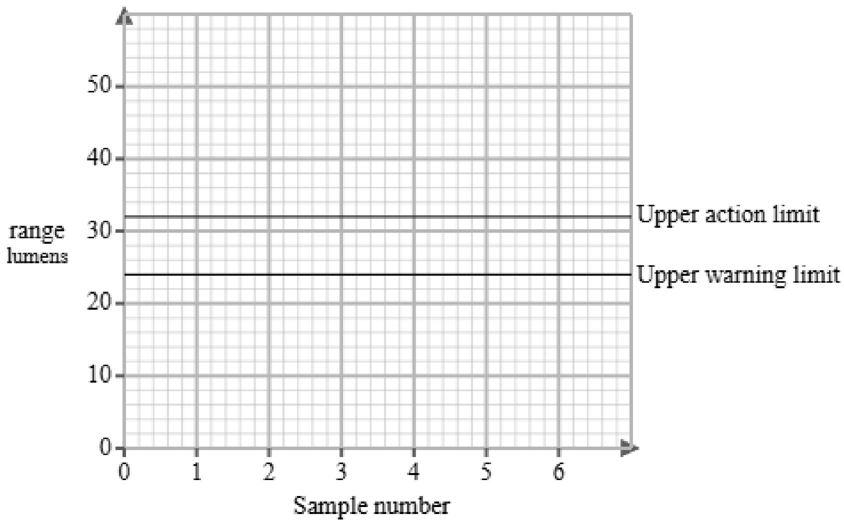
(1 mark)

(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)

15 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) 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)

(b) 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)

16 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)