AQA

Please write clearly in block capitals.

Centre number | | | | | | Candidate number | | | | |

Surname _____

Forename(s) _____

Candidate signature _____

I declare this is my own work.

GCSE
STATISTICS

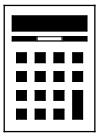
H

Higher Tier Paper 2

Monday 19 June 2023 Afternoon Time allowed: 1 hour 45 minutes

Materials

For this paper you must have:
- a calculator
- mathematical instruments.



Instructions
- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Do not write outside the box around each page or on blank pages.
- If you need extra space for your answer(s), use the lined pages at the end of this book. Write the question number against your answer(s).
- Do all rough work in this book. Cross out any work you do not want to be marked.

Information
- The marks for the questions are shown in brackets.
- The maximum mark for this paper is 80.
- You may ask for more answer paper and graph paper. These must be tagged securely to this answer booklet.

For Examiner's Use	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
TOTAL	

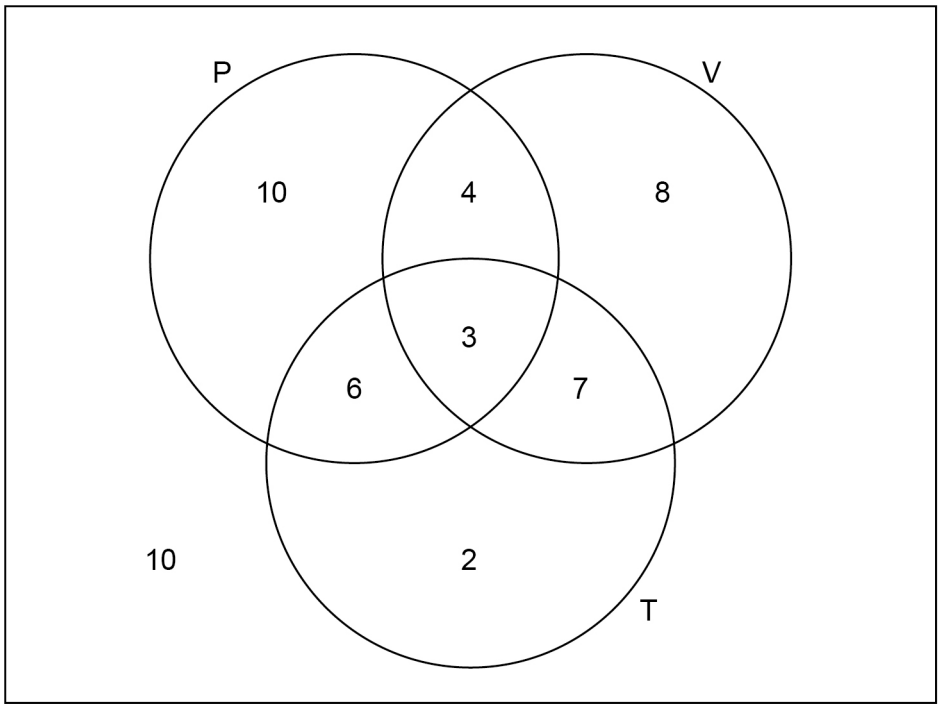
Do not write outside the box

Answer **all** questions in the spaces provided.

1 Tirunesh asks 50 people which of the following instruments they play,

- piano (P)
- violin (V)
- trombone (T).

The results are shown in the diagram.



Tirunesh chooses one person at random.

1 (a) Circle the probability that this person plays the piano.

[1 mark]

$\frac{10}{50}$ $\frac{22}{50}$ $\frac{23}{50}$ $\frac{14}{50}$

1 (b) Circle the value of P(V | T).

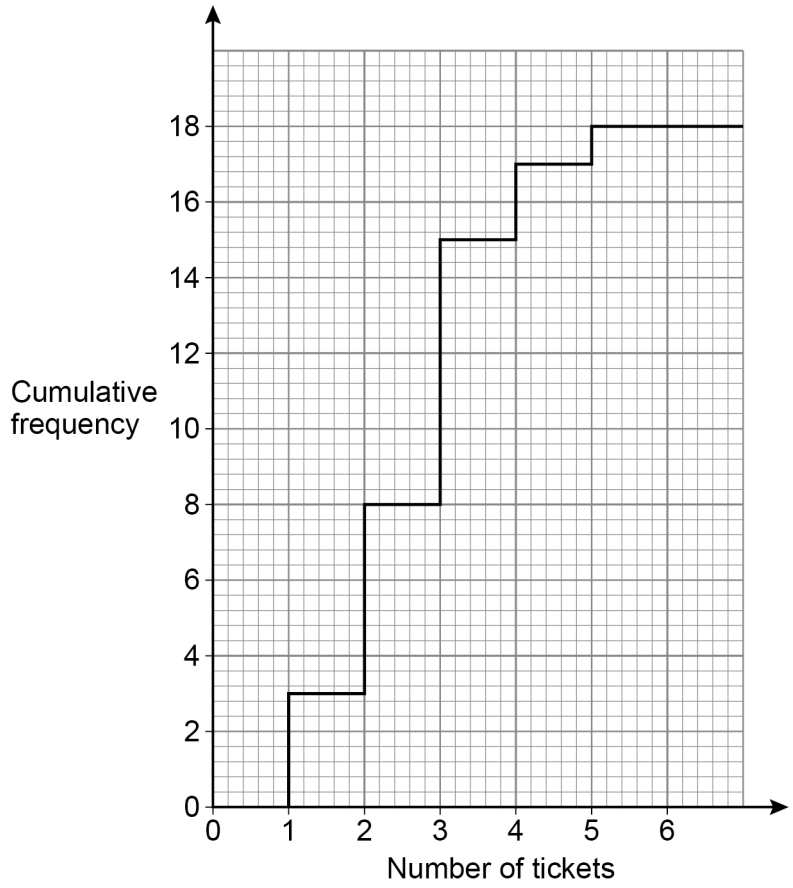
[1 mark]

$\frac{1}{6}$ $\frac{5}{9}$ $\frac{7}{18}$ $\frac{5}{11}$

2

Do not write outside the box

2 The cumulative frequency step polygon represents the number of tickets bought by 18 customers one morning.



Circle the number of customers who bought 3 tickets.

[1 mark]

5 7 8 15

1

3 Tony is investigating students' scores in statistics exams.

He says that it depends on the following factors,

- time spent studying
- score in maths exams.

Circle the term that best describes the type of data Tony uses.

[1 mark]

bivariate multivariate qualitative categorical

1

Turn over ►

Do not write outside the box

4 The table shows information about births in the UK in 2010 and 2019.

	Year	
	2010	**2019**
Number of births	807 300	712 700
Total population	62 260 000	66 650 000
Birth rate		10.7

Source: ONS

4 (a) Complete the table by finding the birth rate in 2010.

Use the formula

$$\text{birth rate} = \frac{\text{number of births}}{\text{total population}} \times 1000$$

[2 marks]

4 (b) The birth rate in Iceland in 2019 was 12.5

Jack concludes,

> "Iceland had a higher number of births than the UK in 2019 because it had a higher birth rate."

Comment on Jack's conclusion.

[1 mark]

3

Do not write outside the box

5 An app allows students to complete work set by their teacher.

Mr Roper wants to test the effectiveness of the app in an experiment.

He tells his class of 30 students that they have a test next week.

To revise, he gives the app to half of the class, whilst the other half just use their books.

Mr Roper ensures that there are the same numbers of boys and girls in each group.

5 (a) Name **one** other consideration for Mr Roper when grouping the students.

[1 mark]

5 (b) The test scores for the **app group** are shown in the stem-and-leaf diagram.

4	5
5	2 3 7
6	4 5 6 8
7	0 3 5 7
8	1 9
9	2

Key: 4 | 5 represents 45%

5 (b) (i) Work out the median score.

[1 mark]

Answer _____ %

5 (b) (ii) Show that the interquartile range is 20%

[2 marks]

Do not write outside the box

5 (c) The percentage scores of the 15 students who were in the **book group** are,

71	46	57	37	50
44	69	40	58	83
42	56	39	55	79

Use these results to complete the back-to-back stem-and-leaf diagram that shows both sets of results on the same diagram.

Remember to complete the key and the labels for the diagram.

[4 marks]

							App group			
						4	5			
						5	2	3	7	
						6	4	5	6	8
						7	0	3	5	7
						8	1	9		
						9	2			

Key | | represents

Do not write outside the box

5 (d) Using your stem-and-leaf diagram and the values from **part (b)**,

compare statistically the scores in the test for those who revised using the app and those who used their books.

[5 marks]

5 (e) Give **one** criticism of the experiment set up by Mr Roper.

[1 mark]

14

Turn over for the next question

Do not write outside the box

6 Frieda, Piran and Elliot are investigating the most popular type of ice cream or ice lolly of the 1200 students at their school.

They want to take a sample of 60 students.

They each suggest a sampling method to collect the data.

Frieda's method

Ask the first 60 students who arrive at school.

Piran's method

Number each student on the register from 1 to 1200

Use a random number generator to find a starting point between 1 and 20

Choose to ask every 20th student after the starting point.

Elliot's method

For each year group, put all the names of the students in a hat.

Pick 5% of the names out for each year group.

Ask each of these students.

6 (a) Write down the name of Frieda's method.

[1 mark]

Answer _____

6 (b) (i) Write down **one** advantage of Piran's method.

[1 mark]

_____

_____

_____

Do not write outside the box

6 (b) (ii) Write down **one** disadvantage of Piran's method.

[1 mark]

6 (c) (i) Write down the name of the sampling technique Elliot suggests.

[1 mark]

Answer _____

6 (c) (ii) The table shows the number of students in each year group at their school.

Year	Number of students
7	260
8	200
9	178
10	280
11	282

Calculate the number of students Elliot should sample from **Year 7**

[2 marks]

Answer _____

Question 6 continues on the next page

Do not write outside the box

6 (d) The table shows the most popular types of ice cream or ice lolly and the percentage of people of different ages and gender who chose them in the UK.

		Gender (%)		Age category (%)			
Ice cream or ice lolly	**All**	**Male**	**Female**	**18–24**	**25–49**	**50–64**	**65+**
Magnum	28	27	28	12	22	39	42
Fab	8	9	8	11	11	6	1
Solero	8	6	9	7	8	8	7
Twister	7	6	7	11	10	2	2
Feast	6	5	7	6	9	6	1
Calippo	5	6	4	7	6	3	2
Orange	3	3	2	1	1	3	10
Cornetto	3	2	3	2	3	2	1
Other	32	36	32	43	30	31	34

Source: adapted from YouGov

Compare the preferences of **different** age groups.

[2 marks]

Comparison 1 _____

_____

_____

Comparison 2 _____

_____

_____

8

Do not write outside the box

Turn over for the next question

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

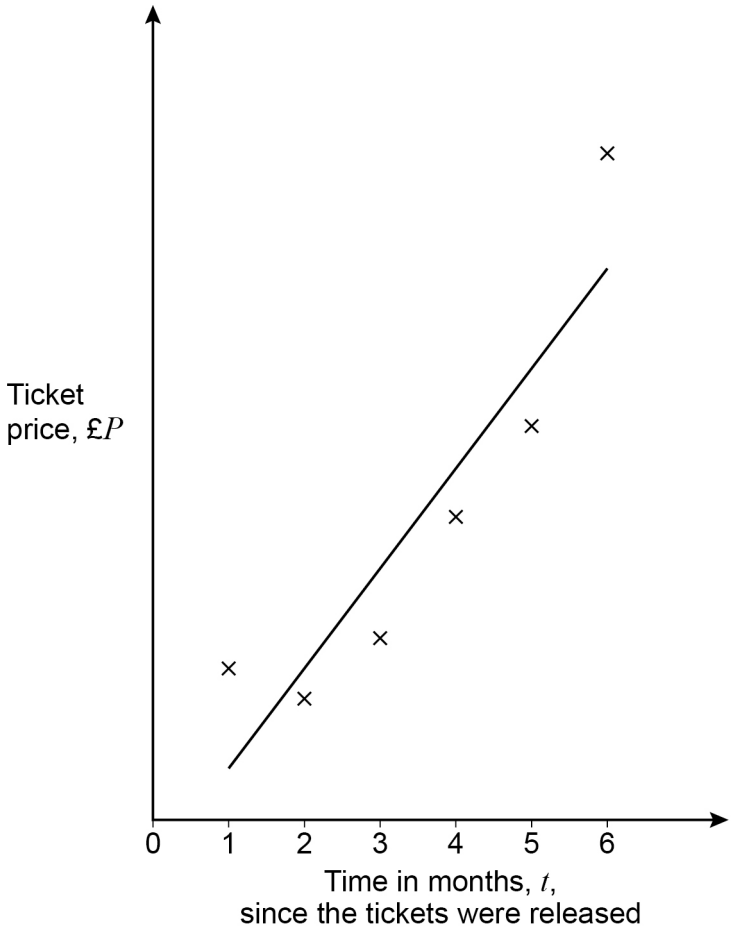
Do not write outside the box

7 Leanne wants to buy a ticket to fly from Manchester to Greece next year.

Tickets are released by the airline exactly **six** months in advance of departure.

She investigates how the price of tickets changes over these six months.

She draws this scatter graph showing the ticket price, £P, plotted against the time in months, t, since the tickets were released.



Leanne calculates that the equation of the regression line for her data is,

$$P = 32t - 16$$

7 (a) Interpret the value of 32 in Leanne's equation.

Give your answer in context.

[1 mark]

Do not write outside the box

7 (b) Leanne does not want to pay more than £100 for her ticket next year.

Use the equation of the regression line to predict the latest time after tickets are released that Leanne could buy a ticket.

[2 marks]

Answer _____ months

7 (c) Give a reason, in context, why the regression equation is **not** appropriate when the tickets are initially released.

[1 mark]

7 (d) The product moment correlation coefficient for Leanne's data is 0.91

Leanne says,

"This shows there is a strong linear relationship."

Looking at the scatter graph, is Leanne's conclusion appropriate?

Tick (✓) a box.

Yes ☐ No ☐

Give a reason for your answer.

[1 mark]

5

Turn over for the next question

Do not write outside the box

8 The table shows average house prices in the UK.

Year	Average price (£)
2016	283 000
2017	280 000

Source: ONS

8 (a) Comment on the reliability of the source of the data.

[1 mark]

8 (b) Using 2016 as the base year, show that the index number for 2017, to 1 decimal place, is 98.9

[1 mark]

8 (c) Using 2016 as the base year, the index number for 2020 is 107.0

8 (c) (i) A house purchased in 2016 cost £430 000

Calculate an estimate for the value of this house in 2020.

[2 marks]

Answer £ _____

Do not write outside the box

8 (c) (ii) Carla bought a house for £215 000 in 2016 and sold it for £233 000 in 2020.

Did Carla's house increase by a **greater** proportion than the index number for 2020 suggests?

Tick (✓) a box.

Yes ☐ No ☐

You **must** show your working.

[2 marks]

8 (d) The table shows some index numbers relating to average house prices for a town.

Year	Index number
2017	100
2019	96.2
2021	104.2

Andy says,

"Average house prices in the town increased by 8% from 2019 to 2021."

Is Andy's conclusion valid?

Tick (✓) a box.

Yes ☐ No ☐

Give a reason for your answer.

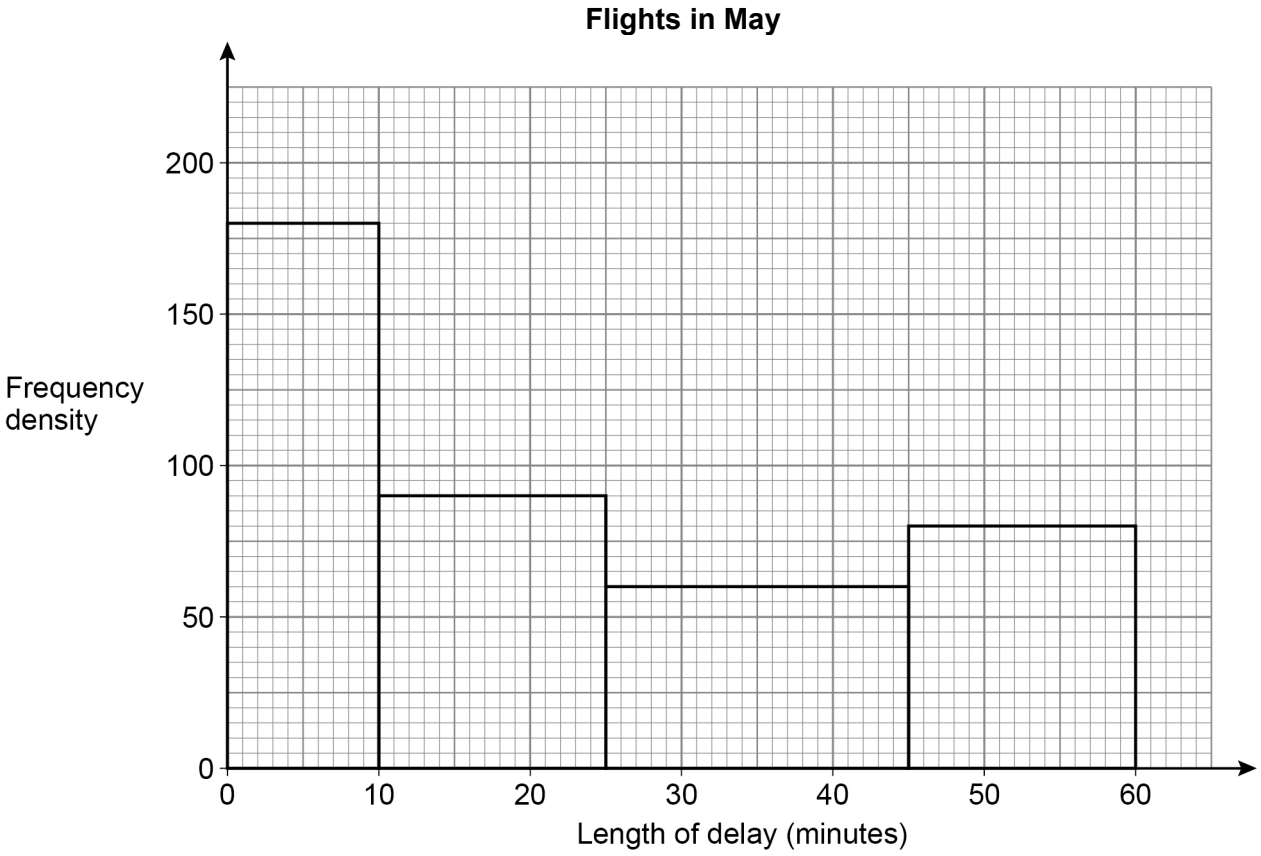
[1 mark]

7

Do not write outside the box

9 An airline records the length of delay, in minutes, whenever a flight is delayed.

The histogram represents this information for flights in **May**.



9 (a) Calculate an estimate for the number of flights in May that were delayed by **more than** 20 minutes.

[3 marks]

Answer _____

9 (b) (i) Give a reason why it is **not** possible to calculate the exact value of the mean length of a delay from the histogram.

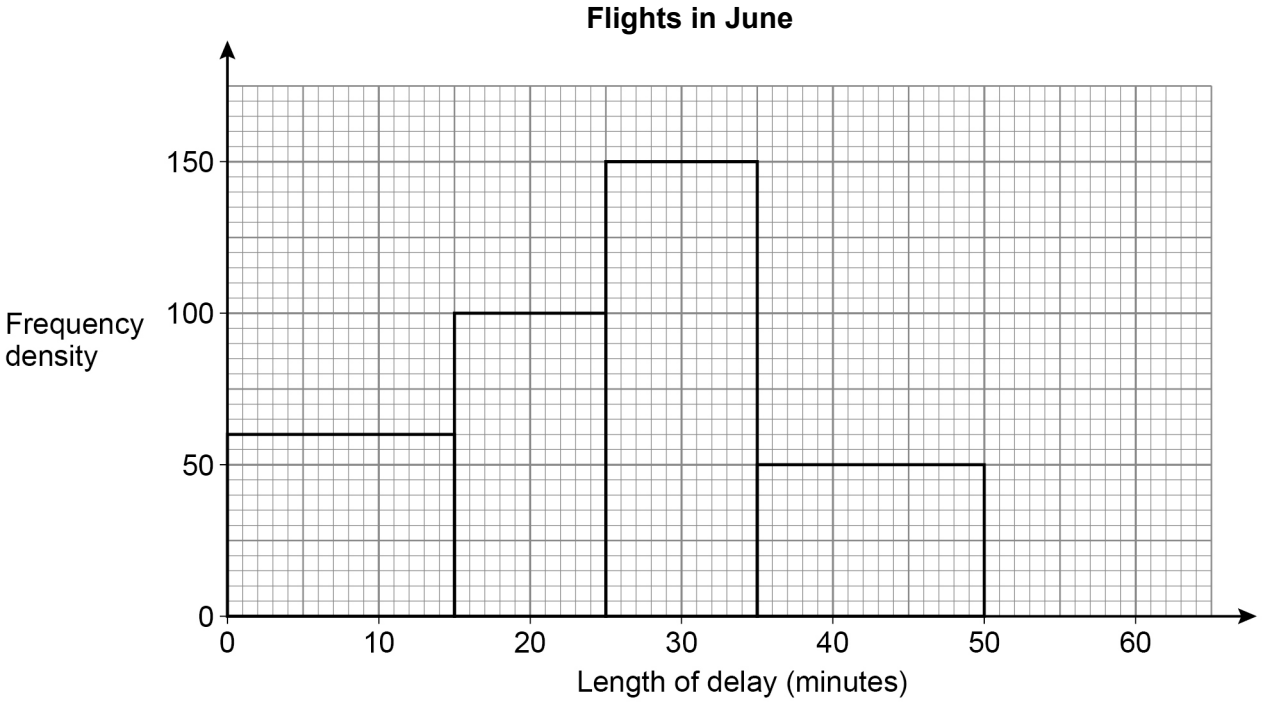
[1 mark]

Do not write outside the box

9 (b) (ii) Give a reason why the mean may **not** be representative for the length of a delay in May.

[1 mark]

9 (c) The histogram represents the length of delay, in minutes, for flights in **June**.



The director of the airline says,

"The range for the length of a delay was **smaller** in June than in May."

Complete these sentences to comment on this statement.

[2 marks]

The director may be correct because _____

The director may **not** be correct because _____

7

Turn over ►

Do not write outside the box

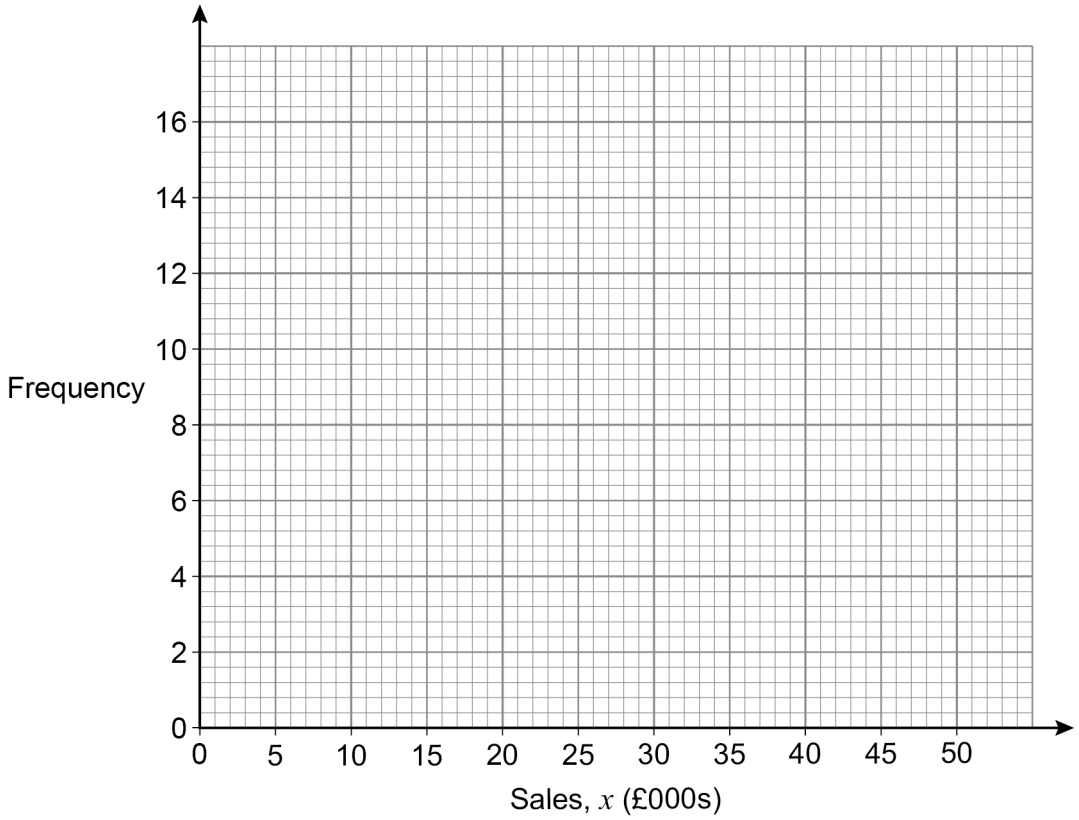
10 Jutta works in the finance department for a retailer.

The table shows information about monthly sales figures over the past 40 months.

Sales, x (£000s)	**Frequency**
$0 \leqslant x < 10$	2
$10 \leqslant x < 20$	5
$20 \leqslant x < 30$	11
$30 \leqslant x < 40$	14
$40 \leqslant x < 50$	8

10 (a) Draw a frequency polygon to represent Jutta's data.

[3 marks]



Do not write outside the box

10 (b) Jutta chooses **two** of the months at random to examine their sales figures.

Calculate the probability that the sales figure for **both** months is at least £20 000

You **must** show your working.

[3 marks]

Answer _____

10 (c) Calculate an estimate of the median monthly sales.

Give your answer to 3 significant figures.

[3 marks]

Answer £ _____

9

Turn over for the next question

Do not write outside the box

11 A triathlon is an athletics event which has three sections,

- a swim
- a cycle
- a run.

11 (a) Owen wants to investigate whether there is a relationship between the times taken for the swim and cycle sections.

Suggest a possible hypothesis for Owen's investigation.

[1 mark]

11 (b) Owen uses some of the results from the 2016 Olympic Games.

The table shows information for 8 different athletes.

Athlete	Swim (min : sec)	Cycle (min : sec)	Rank Swim	Rank Cycle	Difference	
Brownlee	17:24	55:04	1	3	–2	
Schoeman	17:25	55:01	2	1	1	
Murray	18:20	55:35	8	5	3	
Pereira	18:03	55:52	7	6	1	
Van Riel	17:27	55:03	4	2	2	
Mola	17:37	56:18	6	8	–2	
Royle	17:26	55:05	3	4	–1	
Bailie	17:31	56:11	5	7	–2	

Source: triathlon.org

Do not write outside the box

11 (b) (i) Calculate the Spearman's rank correlation coefficient between the times taken to complete the swim and cycle sections.

Use $r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$

You may use the blank column in the table to help you.

[3 marks]

Answer _____

11 (b) (ii) Interpret the value of the answer found in **part (b)(i)** in context.

[1 mark]

Question 11 continues on the next page

Do not write outside the box

11 (c) The values of Spearman's rank correlation coefficient between other sections of the triathlon are shown in the table.

Sections	**Spearman's rank correlation coefficient**
Swim and Run	−0.50
Cycle and Run	−0.26

Owen says,

> "There is a **stronger** relationship between the swim and run sections than between the cycle and run sections."

Is Owen correct?

Tick (✓) a box.

Yes ☐ No ☐

Give a reason for your answer.

[1 mark]

_____

_____

_____

6

Do not write outside the box

12 The weighted mean of two numbers is calculated.

One number has a value of 16 and a weighting of 7

The other number has a value of 11 and a weighting of 3

Circle the value of the weighted mean of the two numbers.

[1 mark]

1

12.5 13.5 14.5 19.7

Turn over for the next question

Do not write outside the box

13 A soft drink is sold in cans which are filled by a machine.

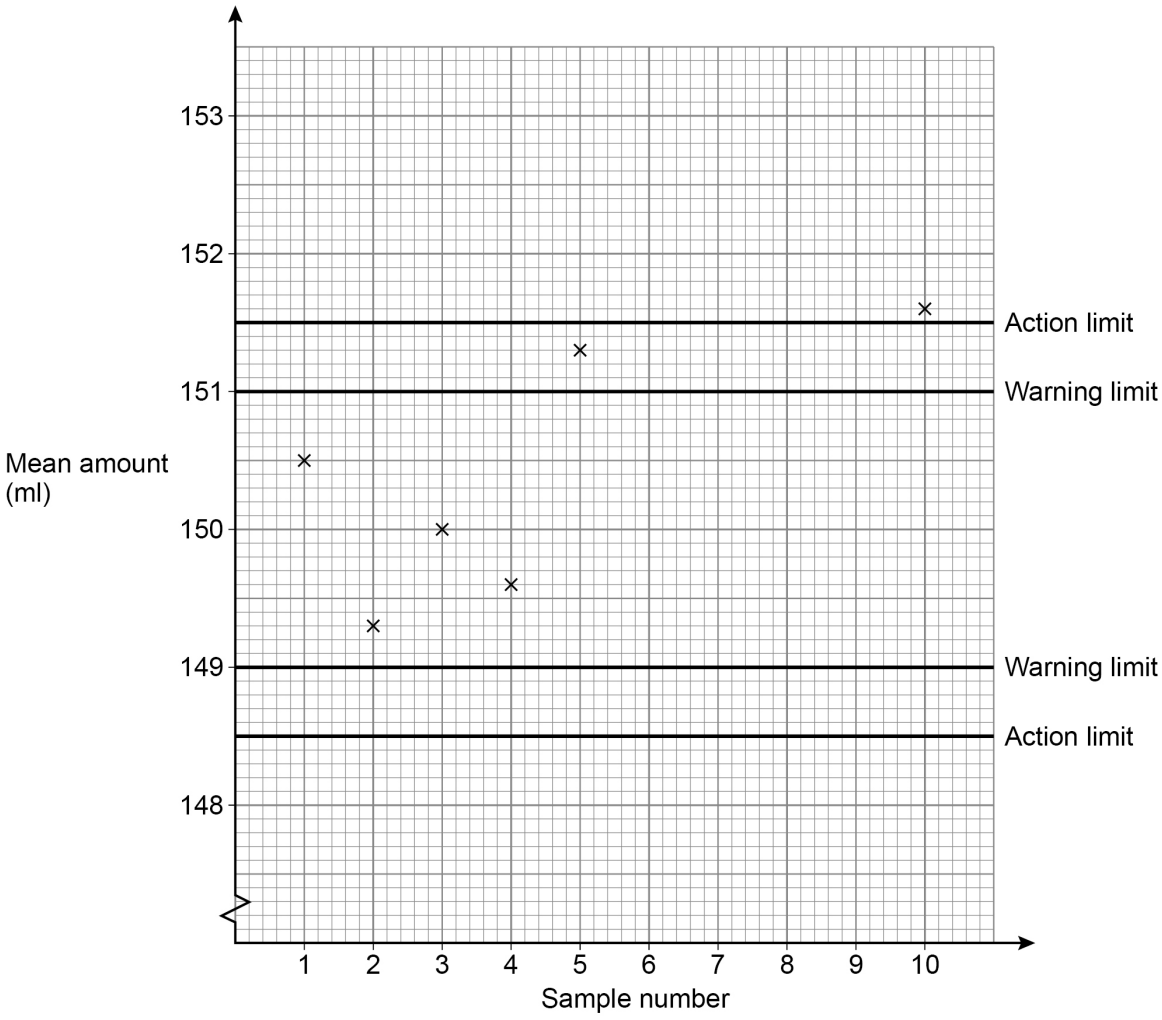
Balasz takes a sample of cans every hour.

He finds the mean amount of drink in each sample.

The table shows the mean of each sample.

Sample number	1	2	3	4	5	6	7	8	9	10
Mean amount (ml)	150.5	149.3	150.0	149.6	151.3	150.8	149.8	151.2	149.3	151.6

The control chart shows some of the sample means and the warning and action limit lines.



Do not write outside the box

13 (a) Complete the control chart.

[2 marks]

13 (b) The machine is set so that,

- the mean amount of drink in a can is x ml
- the standard deviation of the sample mean is y ml

The warning limit lines for the machine are drawn at $x + 2y$ and $x - 2y$

Use the control chart to write down the value of x and work out the value of y

[2 marks]

$x =$ _____

$y =$ _____

13 (c) (i) Balasz actually took Sample 6 immediately after analysing Sample 5

Comment on whether this was appropriate.

Give a reason for your answer.

[1 mark]

_____

_____

_____

13 (c) (ii) Explain what action Balasz should take based on Sample 10

[1 mark]

_____

_____

_____

6

Do not write outside the box

14 In a talent show each contestant is given a mark by the judging panel.

The marks of **all** contestants can be modelled by a normal distribution with a mean of 60 and a standard deviation of 15

Daisy, Helen and Poppy each audition for the talent show.

Their standardised scores are shown in the table.

Contestant	**Standardised score**
Daisy	–0.8
Helen	–2.2
Poppy	1.6

14 (a) To qualify for the next round in the competition a contestant needs a mark of 80 or more.

Show that Poppy qualifies for the next round.

Use $\text{standardised score} = \dfrac{\text{value} - \text{mean}}{\text{standard deviation}}$

[2 marks]

14 (b) (i) How does Daisy's mark compare with Helen's mark?

[1 mark]

Do not write outside the box

14 (b) (ii) How does Daisy's mark compare with the marks of **all** contestants?

[1 mark]

4

Turn over for the next question

Do not write outside the box

15 Almaz wants to estimate the number of frogs in a lake.

- She captures a sample of 380 frogs and tags them.
- She returns the frogs to the lake.
- A week later she collects a second sample of 405 frogs.
- In this second sample 3 of the frogs have tags.

15 (a) Show that an **estimate** of the number of frogs in the lake is approximately 51 000

[2 marks]

Do not write outside the box

15 (b) Two years ago, Almaz estimated that there were 59 000 frogs in the lake.

She concludes that the number of frogs in the lake has decreased by about 13%

Do you agree with Almaz?

Tick (✓) a box.

Yes ☐ No ☐

Show your working.

You should also refer to **three** factors that could affect the accuracy of Almaz's estimate, including any assumptions made.

[4 marks]

Factor 1 _____

Factor 2 _____

Factor 3 _____

6

END OF QUESTIONS

Do not write outside the box

There are no questions printed on this page

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

Do not write outside the box

Question number	**Additional page, if required.** **Write the question numbers in the left-hand margin.**

Do not write outside the box

Question number	**Additional page, if required.** **Write the question numbers in the left-hand margin.**
	Copyright information For confidentiality purposes, all acknowledgements of third-party copyright material are published in a separate booklet. This booklet is published after each live examination series and is available for free download from www.aqa.org.uk. Permission to reproduce all copyright material has been applied for. In some cases, efforts to contact copyright-holders may have been unsuccessful and AQA will be happy to rectify any omissions of acknowledgements. If you have any queries please contact the Copyright Team. Copyright © 2023 AQA and its licensors. All rights reserved.