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 1

Wednesday 5 June 2024 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
1–4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
TOTAL	

Do not write outside the box

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

1 A biased six-sided dice is rolled 50 times.

The results are shown below.

Score	Frequency
1	8
2	11
3	7
4	10
5	9
6	5

Use these results to estimate the probability that the dice rolls a 6

Circle your answer.

[1 mark]

$\frac{1}{6}$ $\frac{5}{6}$ $\frac{5}{50}$ $\frac{45}{50}$

2 A school uses this table to give students a score that represents their behaviour.

Score	Behaviour Description
1	Excellent
2	Good
3	Satisfactory
4	Unacceptable

What type of data is the score?

Circle your answer.

[1 mark]

Bivariate Ordinal Continuous Grouped

Do not write outside the box

3 A data set is shown below.

5 5 6 7 21

The value of 21 is identified as an outlier and removed.

Which of the following measures will **not** decrease?

Circle your answer.

[1 mark]

Mean Mode Median Range

4 Each value in a data set is different.

Which of the following measures of spread has the largest value?

Tick (✓) a box.

[1 mark]

The interquartile range. ☐

The interdecile range between the 1st and 9th deciles. ☐

The interpercentile range between the 20th and 80th percentiles. ☐

The range. ☐

4

Turn over for the next question

Do not write outside the box

5 All Year 7 students in two schools take the same French test.

Sanjit wants to know if students in school A or school B score better.

5 (a) Write down a suitable hypothesis that Sanjit could use.

[1 mark]

5 (b) Describe fully the population for Sanjit's investigation.

[1 mark]

Do not write outside the box

5 (c) Sanjit considers the following two methods for collecting scores from a sample of 19 students in school A.

Method A

Ask the first 19 Year 7 students who arrive in the playground.

Method B

Give each student in Year 7 a unique number.

Generate 19 different random numbers.

Use the 19 students whose numbers match the ones generated.

State the name of each method and give one advantage of that method.

[4 marks]

Method A

Name _____

Advantage _____

_____

Method B

Name _____

Advantage _____

_____

Question 5 continues on the next page

Do not write outside the box

5 (d) Sanjit collects the test scores from a sample of 19 Year 7 students in **school A**.

Here are his data.

5	8	9	9	10	11	13	14	14	15
18	19	23	24	26	31	35	37	42	

Draw a box plot to represent these data.

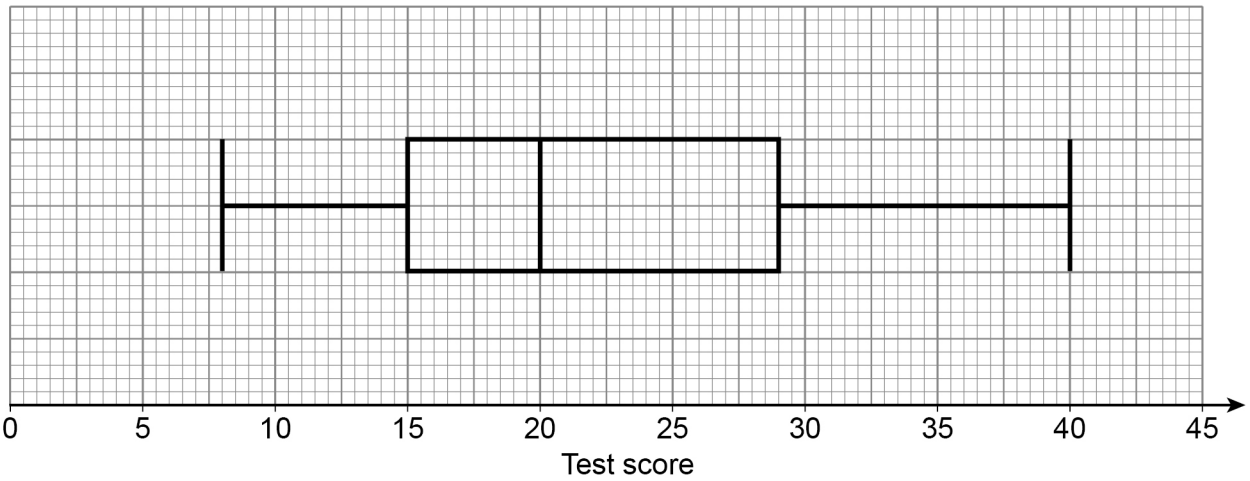
You may use the table to help.

[6 marks]

Lower Quartile	
Median	
Upper Quartile	

Do not write outside the box

Sanjit draws this box plot to show the test scores for his sample of 19 students from **school B**.



5 (e) (i) Use the box plot to write down the median test score for **school B**.

[1 mark]

Answer _____

5 (e) (ii) Compare statistically the median scores for **school A** and **school B**.

[1 mark]

_____

_____

_____

5 (f) (i) Use the box plot to calculate the interquartile range for **school B**.

[1 mark]

_____

_____

Answer _____

5 (f) (ii) Compare statistically the interquartile ranges for **school A** and **school B**.

[1 mark]

_____

_____

_____

16

Turn over ►

Do not write outside the box

6 Alex asks a group of friends how many times they exercised last month.

Here are the data.

1 1 1 2 3 3 5 7 14 30

Alex correctly calculates these three averages.

Mode	1
Median	3
Mean	6.7

Alex wants to know the average that **best** represents these data.

Comment on the suitability of each average.

[3 marks]

Mode _____

_____

_____

Median _____

_____

_____

Mean _____

_____

_____

3

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 A random number generator is programmed to only generate values of 1, 2, 3, 4 and 5

Numbers are generated in groups and recorded to see if the random number generator is fair.

The table shows information about the number of 1s generated.

Group	Number of results generated (cumulative)	Number of 1s recorded (cumulative)
1	10	2
2	20	3
3	30	4
4	40	5
5	50	5
6	60	6
7	70	7
8	80	9
9	90	10
10	100	11

7 (a) Which group contained zero 1s?

[1 mark]

Answer _____

7 (b) (i) Aalim says

"We only need the first set of results because they show the random number generator is fair."

Chloe says

"It is better to use all the results to see if the random number generator is fair."

Give **one** reason why Chloe is correct.

[1 mark]

_____

_____

Do not write outside the box

7 (b) (ii) Use the table to decide if the random number generator is fair.

Tick (✓) a box.

Yes ☐ No ☐ Cannot tell ☐

Give a reason for your answer.

[1 mark]

_____

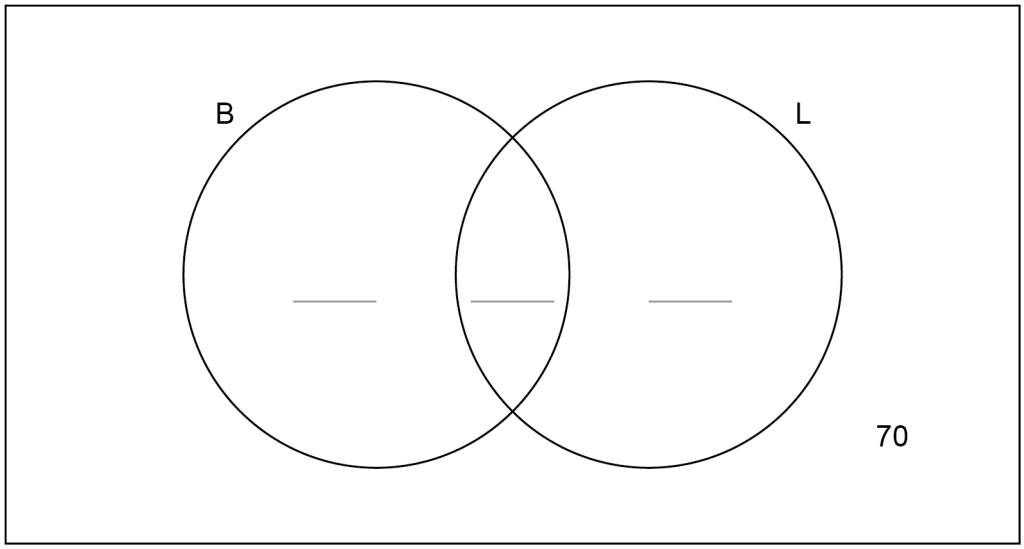
_____

3

Turn over for the next question

Do not write outside the box

8 The Venn diagram shows some information about the 350 photographs in an exhibition.



In the diagram,

B represents the black-and-white photographs

L represents the landscape photographs.

8 (a) Complete the diagram to show that,

- there are 166 black-and-white photographs
- there are 154 landscape photographs.

[3 marks]

Do not write outside the box

8 (b) A photograph is chosen at random.

8 (b) (i) What is the probability that the photograph is not black-and-white and not landscape?

[1 mark]

Answer _____

8 (b) (ii) Given that the photograph is black-and-white, what is the probability that it is **not** landscape?

[2 marks]

_____

_____

Answer _____

6

Turn over for the next question

Do not write outside the box

9 A researcher is investigating the heights of children in their area.

They have access to the medical records of 1200 children.

In the records, each child is categorised into one of four age groups,

- babies
- infants
- juniors
- seniors.

9 (a) Kim suggests they take a sample of 120 children by choosing 30 from each age group.

Give **one** reason why choosing 30 from each group may not be appropriate.

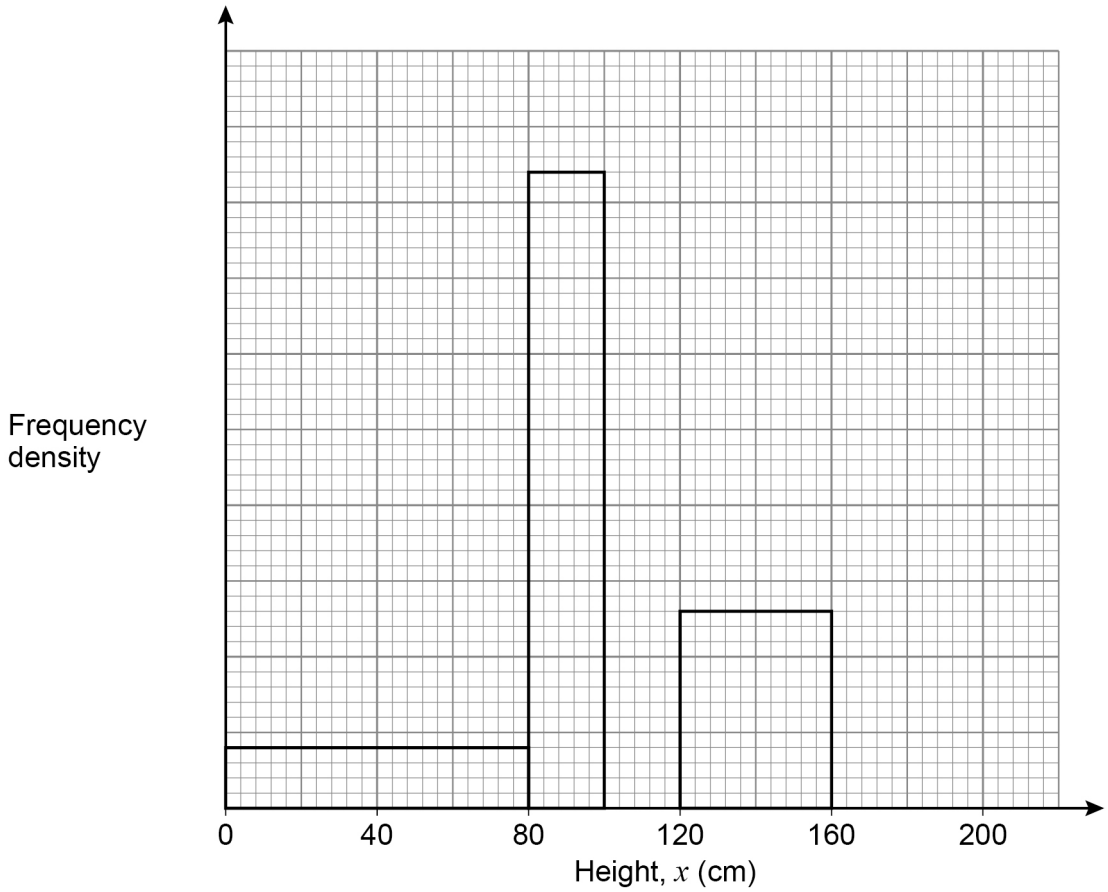
[1 mark]

9 (b) The researcher uses a different method to take a sample of 120 children.

Some information about the heights of the children in the sample is shown in the table and histogram.

Height, x (cm)	**Frequency**
$0 < x \leqslant 80$	
$80 < x \leqslant 100$	42
$100 < x \leqslant 120$	26
$120 < x \leqslant 160$	
$160 < x \leqslant 180$	10

Do not write outside the box



Complete the table and histogram.

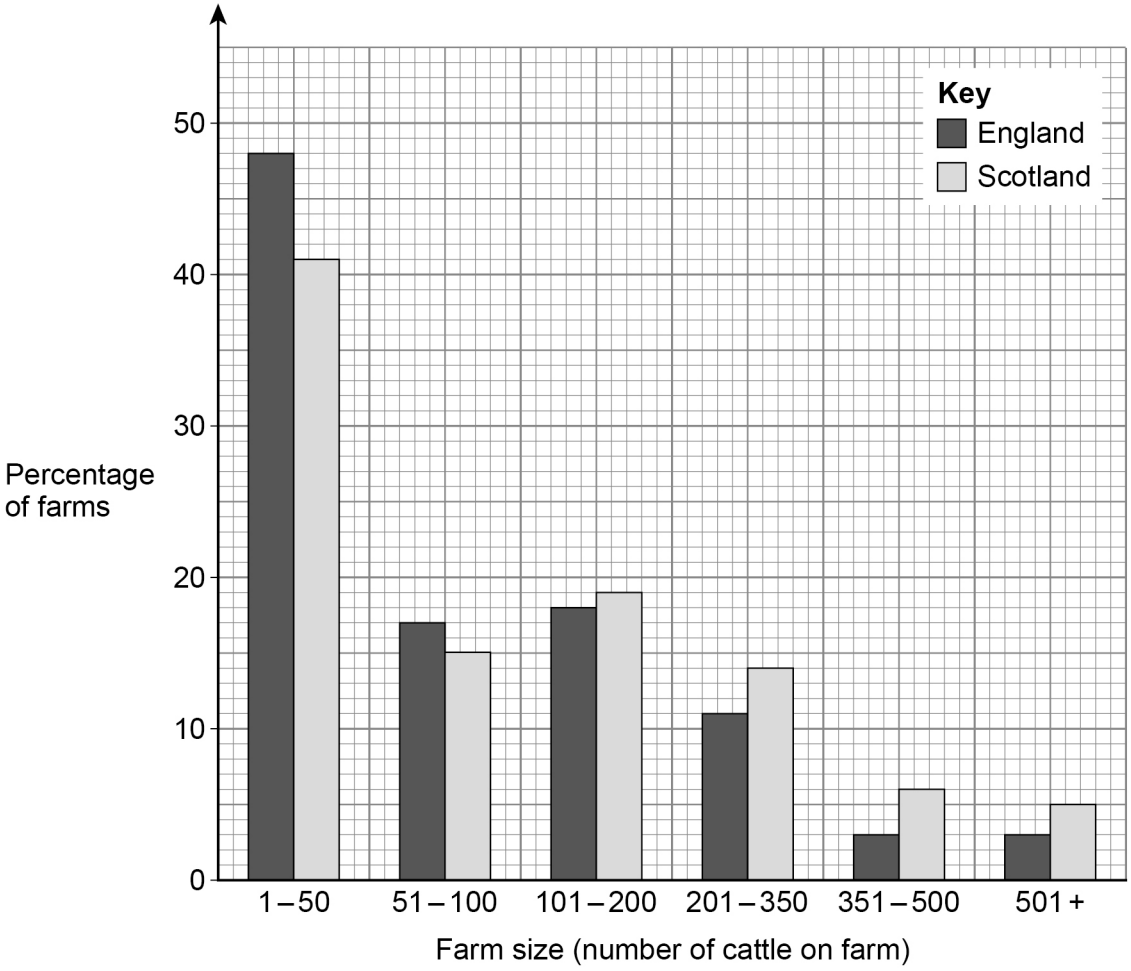
[5 marks]

6

Turn over for the next question

Do not write outside the box

10 The dual bar chart shows information about the number of cattle on farms in England and Scotland.



Source: Adapted from Gov.uk

10 (a) Is the 90th percentile for **Scotland** between 201 and 350 cattle?

Tick (✓) a box.

Yes ☐ No ☐

Justify your answer.

[2 marks]

Do not write outside the box

10 (b) The following statement is made,

> "Scotland has a greater proportion of large farms, so there are more cattle in Scotland than in England."

Give two reasons why this may **not** be correct.

[2 marks]

1 _____

_____

_____

2 _____

_____

_____

10 (c) Describe the skew for the data about farms in England.

[1 mark]

_____

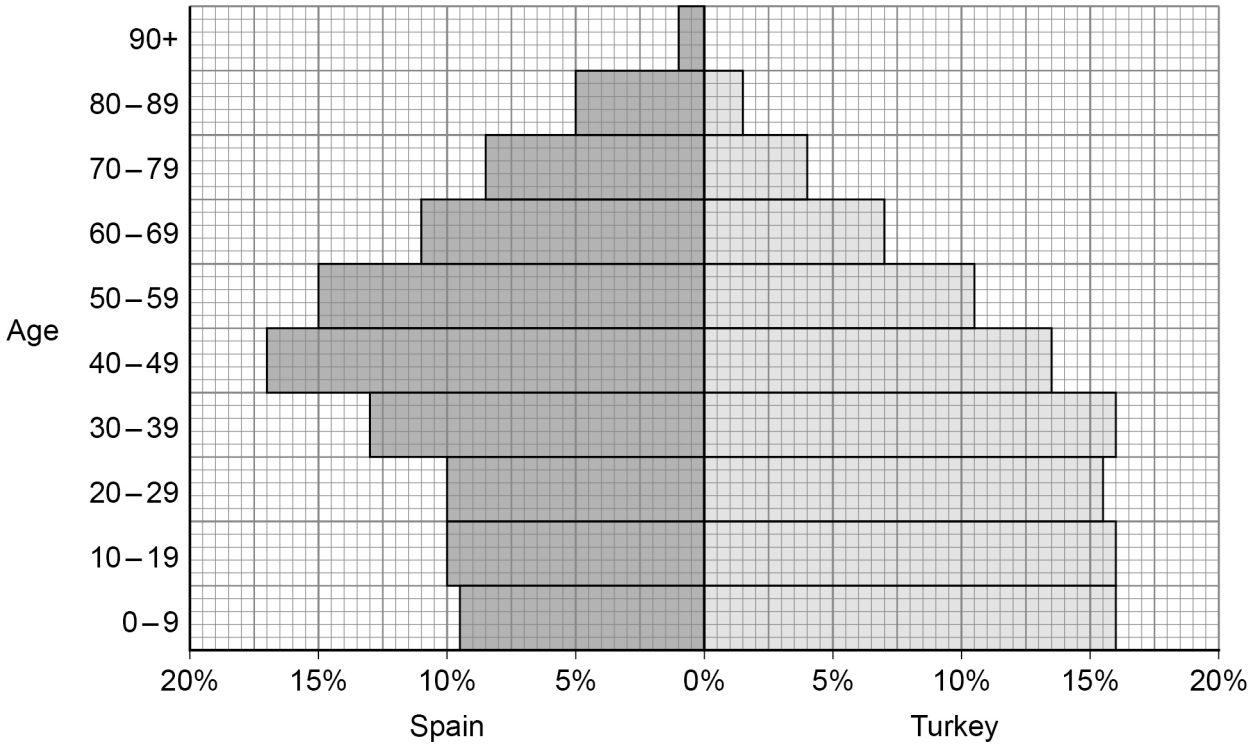
_____

5

Turn over for the next question

Do not write outside the box

11 Aiyla has this diagram comparing the ages of the populations in Spain and Turkey in 2019.



Source: Populationpyramid.net

11 (a) (i) Use the diagram to work out the percentages of the populations aged below 40 in Spain and in Turkey.

[3 marks]

Spain _____ %

Turkey _____ %

Do not write outside the box

11 (a) (ii) Hence compare the percentages of the populations aged below 40 in Spain and in Turkey.

[1 mark]

11 (b) Aiyla finds this information.

Country	Average life expectancy (years)
Spain	83.5
Turkey	77.7

Write down one feature of the graph that supports this information.

[1 mark]

5

Turn over for the next question

Do not write outside the box

12 The table shows the population of England and Wales in 2017 and 2020.

It also shows the birth rate in 2017.

Year	Population for England and Wales	Birth rate (per 1000 population)
2017	58 744 600	11.6
2020	59 719 700	

Source: ONS

The number of births in 2020 was 9.6% lower than the number of births in 2017.

Complete the table using the formula

$$\text{Birth rate (per 1000 population)} = \frac{\text{number of births}}{\text{population}} \times 1000$$

[4 marks]

4

Do not write outside the box

13 A race has a running section and a cycling section.

Seb takes part in the race.

The table shows

- Seb's time for each section
- the mean time and standard deviation of the times for all of the racers for each section.

Section	Seb's time (minutes)	Mean time of all racers (minutes)	Standard deviation of all racers (minutes)
Running	41.8	42.7	4.4
Cycling	68.8	70.7	11.7

Seb's coach says he did better in the running section than in the cycling section.

Is Seb's coach correct?

Tick (✓) a box.

Yes ☐ No ☐

You **must** show your working.

Use the formula

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

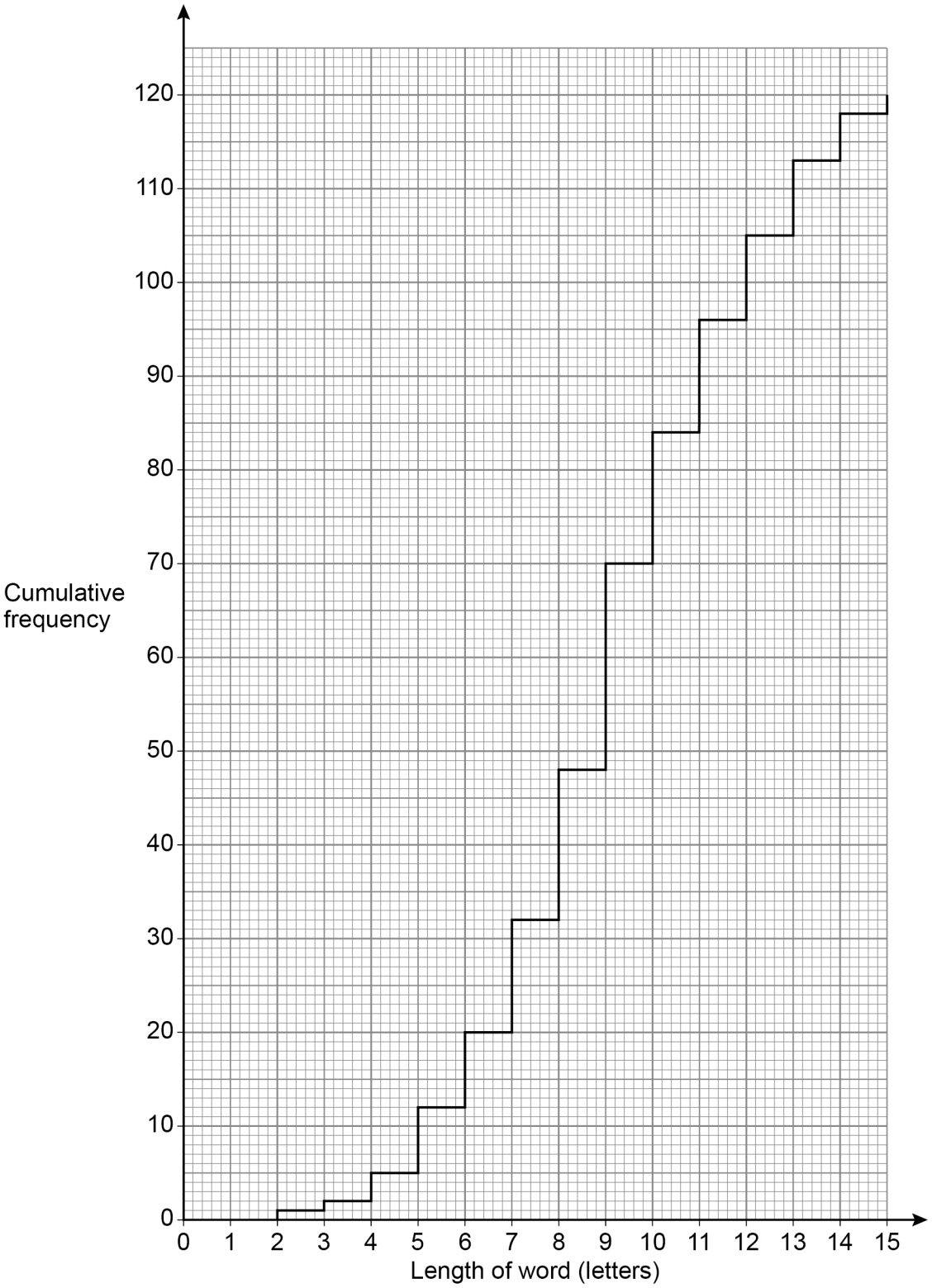
[3 marks]

3

Do not write outside the box

14 Charlie is investigating the lengths of words used in crosswords.

The cumulative frequency step polygon shows the lengths of 120 words used in crosswords in his local newspaper.



Do not write outside the box

14 (a) Work out the percentage of words which had more than 8 letters.

[2 marks]

Answer _____ %

14 (b) (i) Use the graph to complete the table.

[2 marks]

Lower Quartile	Upper Quartile	Interquartile Range

14 (b) (ii) Using the values in your table, determine if there are any outliers.

You **must** show your working.

Tick (✓) a box.

There **are** outliers ☐ There are **no** outliers ☐

[3 marks]

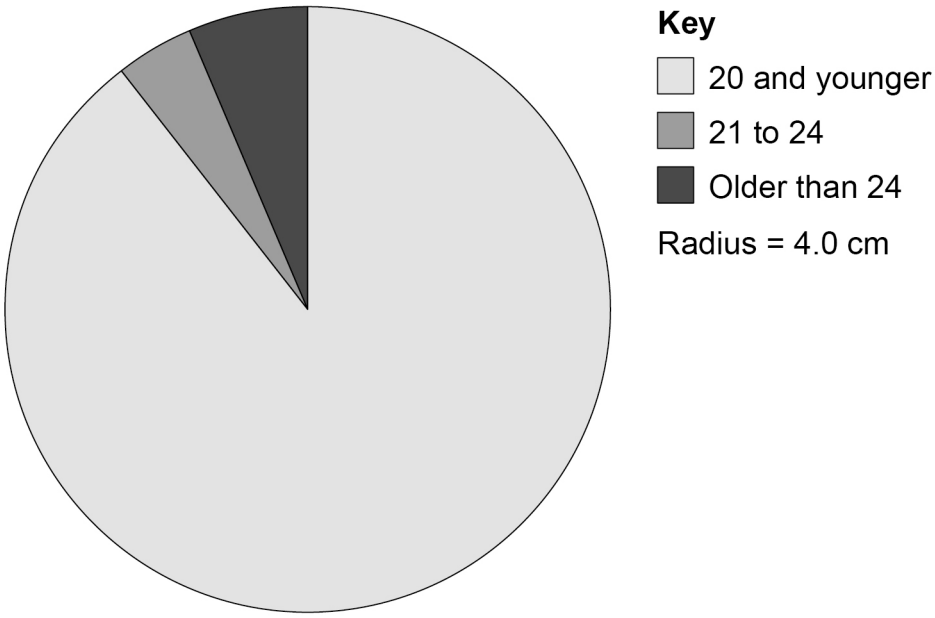
7

Turn over for the next question

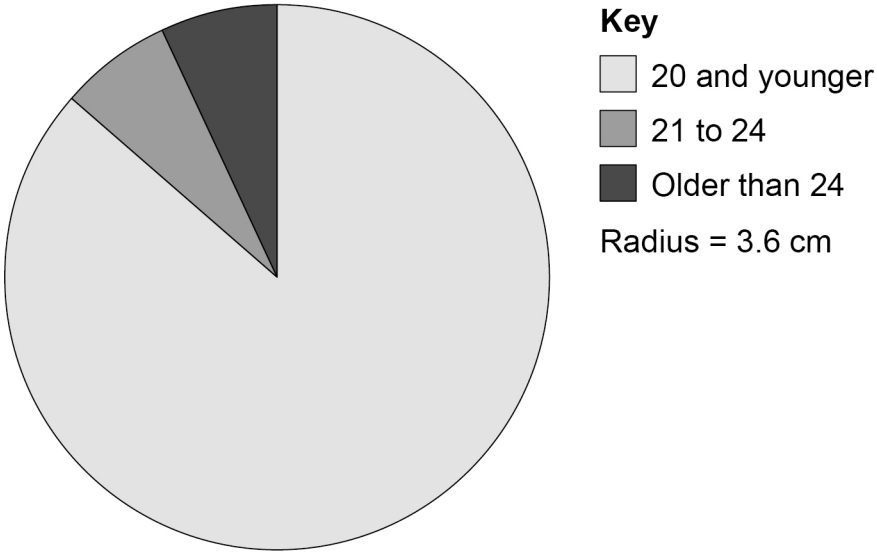
Do not write outside the box

15 The comparative pie charts show information about the ages (in years) of students in two countries who finished university in 2021.

Country A



Country B



Do not write outside the box

15 (a) More students finished university in 2021 in country A than in country B.

How do the pie charts show this?

[1 mark]

15 (b) 618 000 students finished university in 2021 in country B.

Calculate the number of students who finished university in 2021 from country A that were older than 24

You **must** show your working.

[5 marks]

Answer _____

6

Turn over for the next question

Do not write outside the box

16 (a) Emily wants to estimate the population of honey badgers in a large area.

She uses the following method,

- capture 5 honey badgers
- mark them using coloured chalk dust and release them
- three hours later capture 50 honey badgers.

Write down **three** problems with this method.

[3 marks]

1 _____

_____

_____

2 _____

_____

_____

3 _____

_____

_____

Do not write outside the box

16 (b) Roberta wants to estimate the number of honey badgers in a different area.

She uses this method.

1st sample	**2nd sample**
40 captured and then marked	35 captured 8 were marked

Use this information to estimate the number of honey badgers in Roberta's area.

[2 marks]

Answer _____

5

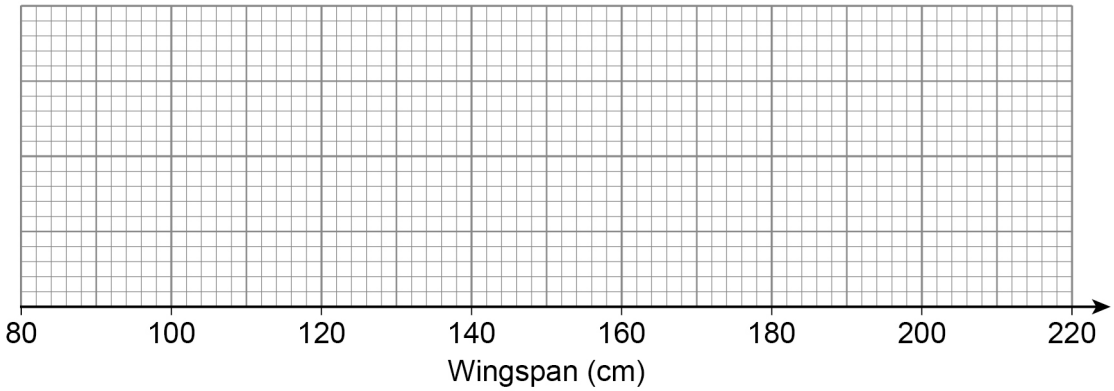
Turn over for the next question

Do not write outside the box

17 The wingspan of the Canada goose is normally distributed with a mean of 156 cm and a standard deviation of 10 cm.

17 (a) Draw a sketch of this distribution on the grid.

[3 marks]



17 (b) The wingspan of the snow goose is normally distributed with a mean of 148 cm and a standard deviation of 6 cm.

A snow goose is selected at random.

Work out the probability that it has a wingspan between 136 cm and 154 cm.

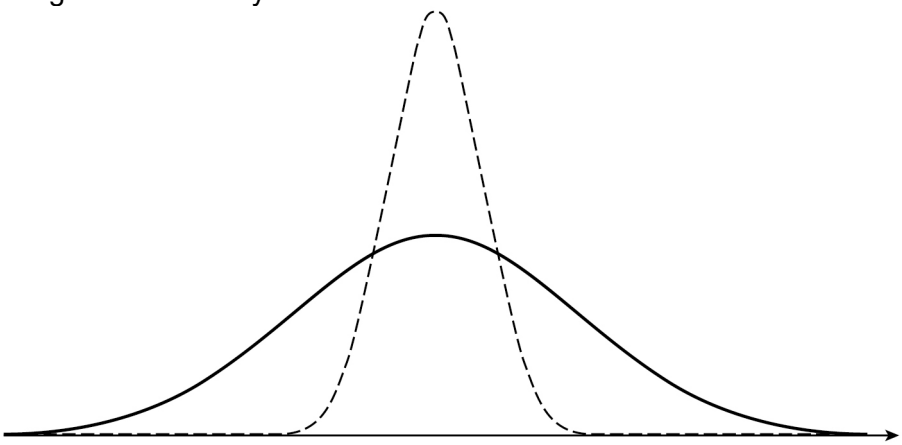
[3 marks]

Answer _____

Do not write outside the box

17 (c) This diagram shows some information about the wingspan of the greylag goose.

The diagram has no key or scale.



The solid line shows the distribution of the wingspan of the greylag goose.

Give **one** reason why the dashed line could show the distribution of the sample means of the wingspan of the greylag goose.

[1 mark]

_____

_____

_____

7

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 © 2024 AQA and its licensors. All rights reserved.