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

Thursday 11 June 2020 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 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	
16	
TOTAL	

Do not write outside the box

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

1 The table shows the index numbers for the cost of an item in different years.

Year	2016	2017	2018	2019
Index number	95	100	90	115

Circle the base year.

[1 mark]

2016 2017 2018 2019

1

2 Here are some summary measures for a distribution.

Smallest value	**2nd decile**	**Largest value**
11	35	161

Range = Largest value − Smallest value

The difference between the 2nd and 8th deciles is 30% less than the range.

Circle the value of the 8th decile.

[1 mark]

80 105 140 155

8^{th} decile − 2^{nd} decile = Range − $\left(\frac{30}{100} \times \text{Range}\right)$

8^{th} decile − 35 = $(161 - 11) - \left(\frac{30}{100} \times (161-11)\right)$

8^{th} decile = 35 + 150 − 45 = 140

1

3 The geometric mean of 3 and x is 6

Circle the value of x.

[1 mark]

2 4 9 12

geometric Mean = $\sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdots X_n}$

where n is the number of terms that are multiplied

$6 = \sqrt{3 \cdot x}$

$6^2 = 3x$

$36 = 3x$

$\Rightarrow x = \frac{36}{3} = 12$

1

4 Here are four scatter diagrams.

Diagram A

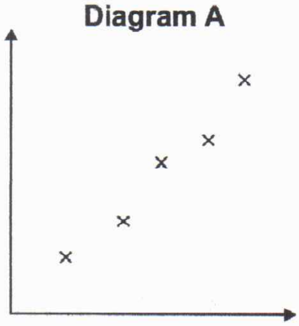


Diagram B

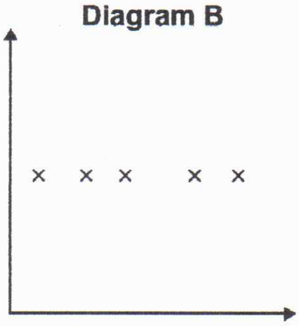


Diagram C

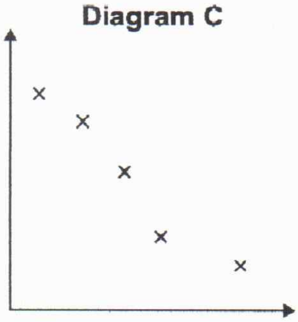
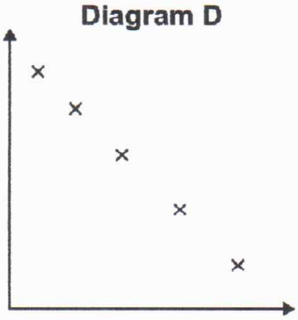


Diagram D



Circle the letter of the scatter diagram for which the Pearson's product moment correlation coefficient is −1

[1 mark]

A B C D

1

Turn over for the next question

Do not write outside the box

5 Marcus is planning a Driver Safety course.

He wants to give the people attending the course a questionnaire to complete.

5 (a) Marcus wants to know how far each person usually drives in a week.

Write a closed question that Marcus could ask to find out this information.

Include a response section.

[3 marks]

What distance in Kilometres do you usually drive each week?

0 – 20   21 – 40   41 – 60   more than 60

5 (b) Marcus also wants to know whether people regularly drive faster than the speed limit.

He plans to collect the information using this method.

> He asks each person to secretly throw a dice.
> The person then answers as follows:
> - if the person gets an odd number, they answer 'Yes'
> - if the person gets an even number, they truthfully answer the question,
>   'Do you regularly drive faster than the speed limit?'

5 (b) (i) Why does Marcus use this method?

[1 mark]

To improve the response rate to the question

Do not write outside the box

5 (b) (ii) Marcus collects data from 100 people using this method.

60 people give the answer 'Yes'.

Marcus says,

"60% of these people regularly drive faster than the speed limit."

Explain why Marcus is wrong.

[1 mark]

some people said yes because they threw an odd number

5

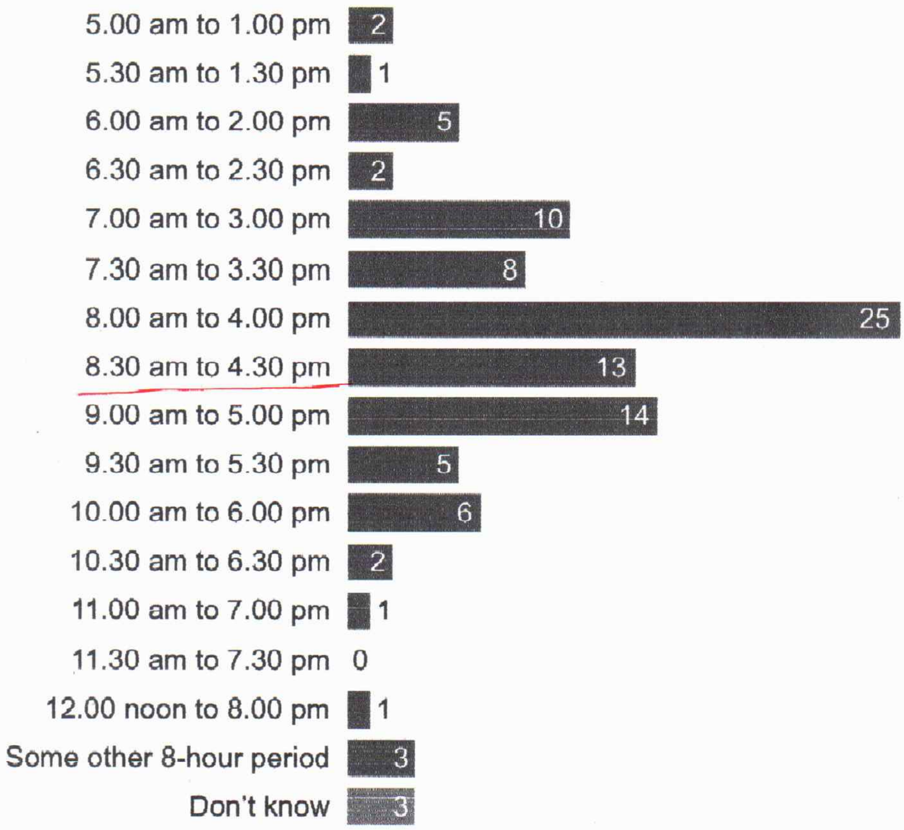
Turn over for the next question

Do not write outside the box

6 A YouGov survey was carried out with nearly 2000 British working adults who have an 8-hour working day.

They were asked which period of 8 hours they would prefer to work.

YouGov produced this summary graph showing the **percentage** of each response, rounded to the nearest whole number.



Source: yougov.com

6 (a) Show that about two-thirds of adults questioned wanted to work **earlier** than the traditional 9 am to 5 pm working hours.

[2 marks]

2+1+5+2+10+8+25+13

2+1+5+2+10+8+25+13+14+5+6+2+1+1+3+3

$= \dfrac{66}{100}$

which is about $\frac{2}{3}$

Do not write outside the box

6 (b) Amber says,

"**None** of the adults questioned wanted to start work at 11.30 am."

Is Amber correct?

Tick (✓) a box.

Yes ☐ No ☐ Cannot tell ☑

Give a reason for your answer.

[1 mark]

Some of the people who gave a response of 'don't know' might want to start work at 11:30 am.

6 (c) Give **one** reason why these results will **not** apply to all British working adults.

[1 mark]

Not all British working adults have fixed working hours that is 8 hours a day

4

Turn over for the next question

Do not write outside the box

7 200 students, 200 parents with young children and 200 retired people were asked what was the first thing they did on their mobile phones that day.

The results are shown in the table.

	Social media	Gaming	News	Other
Students	124	52	13	11
Parents	120	8	37	35
Retired	88	11	67	34
	Σ 332	Σ 71	Σ 117	Σ 80

7 (a) One of the people is chosen at random.

7 (a) (i) Work out the probability that this person goes on social media first that day.

[2 marks]

$P(\text{social media}) = \dfrac{124+120+88}{200+200+200}$

$= \dfrac{332}{600} = 0.5533$

Answer 0.5533

7 (a) (ii) Work out the probability that this person does **not** go on gaming first that day.

[2 marks]

$P\left(\text{does not go on gaming}\right) = \dfrac{P(\text{goes on social media}) + P(\text{goes to news}) + P(\text{other})}{\text{Total number of people}}$

$= \dfrac{332+117+80}{600} = \dfrac{529}{600} = 0.8817$

Answer 0.8817

7 (b) One of the people who went on gaming first that day is chosen at random.

What is the probability that this person is retired?

[2 marks]

$P(\text{retired}) = \dfrac{11}{71} = 0.1549$

Answer 0.1549

Do not write outside the box

7 (c) Work out the probability that **two** of the 200 retired people, chosen at random, both went on news first that day.

Give your answer to three decimal places.

[3 marks]

P(1 person going) = $\frac{67}{200}$

~~P(second person going) = $\frac{66}{199}$~~

P(2 of retired people going on news) = $\frac{67}{200} \times \frac{66}{199}$ = 0.1111

Answer 0.1111

7 (d) Joe looks at the data in the table and makes the two statements below.

Is each statement correct?

Give a reason for each decision.

[2 marks]

Statement 1 Most of these 600 people went on social media first **that** day.

Tick (✓) a box.

Yes ✓ No ☐ Cannot tell ☐

Reason More than half of the people went on social media first that day

Statement 2 Most of these 200 retired people go on social media first **every** day.

Tick (✓) a box.

Yes ☐ No ☐ Cannot tell ✓

Reason The results from the data given are just for one day and might not be the same every day.

11

Do not write outside the box

8 A deadly disease currently has no treatment.

A researcher develops a drug which she believes will treat the disease.

She suggests a statistical experiment to test her drug.

Infect six people chosen at random with the disease.
Give the drug to all six people.
Record whether each person recovers or not.

8 (a) Write down **two** problems with the researcher's experiment.

[2 marks]

Problem 1 The sample size is too small

. The results will not be reliable with just

6 people

Problem 2 The researcher cannot infect randomly

chosen people with a deadly disease.

8 (b) The researcher carries out a more suitable experiment.

She writes an article for a magazine to highlight her results.

She gives the name of each patient in the experiment and records how they responded to the drug.

The magazine editor asks the researcher to rewrite her article.

Explain why.

[1 mark]

Names of patients should not be included in the

experiment, names should be confidential

3

Do not write outside the box

9 In an experiment, Paulo throws three fair coins.

He repeats the experiment 120 times.

How many times should he expect to throw three heads?

[2 marks]

P(throwing three heads in the 1st throw) $= \frac{1}{2} \times \frac{1}{2} \cdot \frac{1}{2} = 1/8$

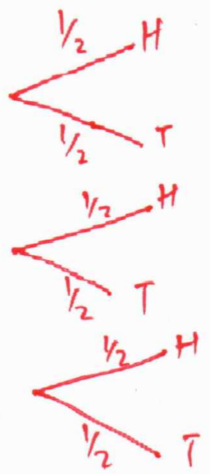
Number of times $= \frac{1}{8} \times$ ~~total~~ number of times the coins are thrown

$= 1/8 \times 120 = 15$

Answer 15

2

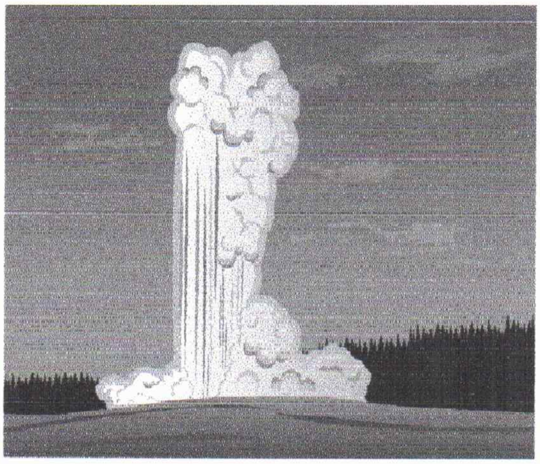
Turn over for the next question



$P(HHH) = 1/2 \times 1/2 \times 1/2 = 1/8$

Do not write outside the box

10 A geyser is a spring which erupts from time to time and shoots a column of hot water into the air.



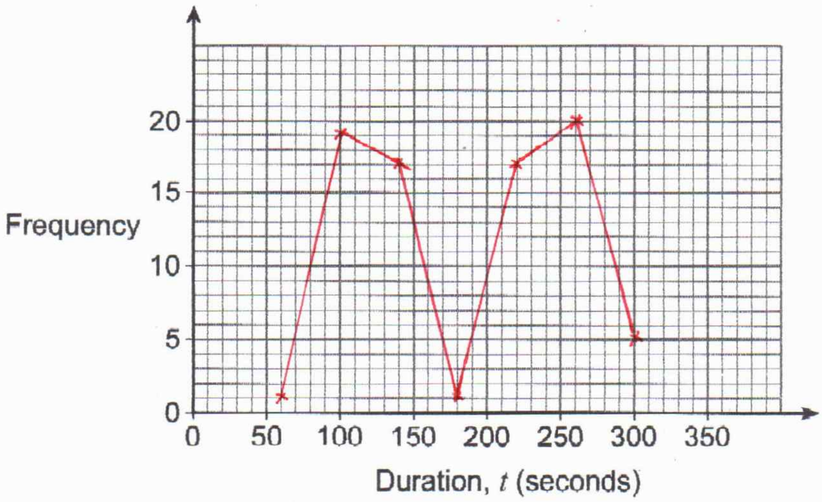
The table shows the duration of 80 eruptions of a geyser.

Mid-point	**Duration, t (seconds)**	**Frequency**
60	$40 < t \leqslant 80$	1
100	$80 < t \leqslant 120$	19
140	$120 < t \leqslant 160$	17
180	$160 < t \leqslant 200$	1
220	$200 < t \leqslant 240$	17
260	$240 < t \leqslant 280$	20
300	$280 < t \leqslant 320$	5
	TOTAL	**80**

Do not write outside the box

10 (a) Draw a frequency polygon to show this information.

[3 marks]



10 (b) Calculate an estimate of the mean duration of an eruption.

Use $\sum ft = 14\,960$

[1 mark]

Mean $= \frac{\Sigma ft}{\Sigma f} = \frac{14960}{80} = 187$

Answer 187 seconds

10 (c) Give a reason why the mean is **not** a typical value for this set of data.

[1 mark]

The mean falls in the interval $160 < t \leq 200$ which has a frequency of 1

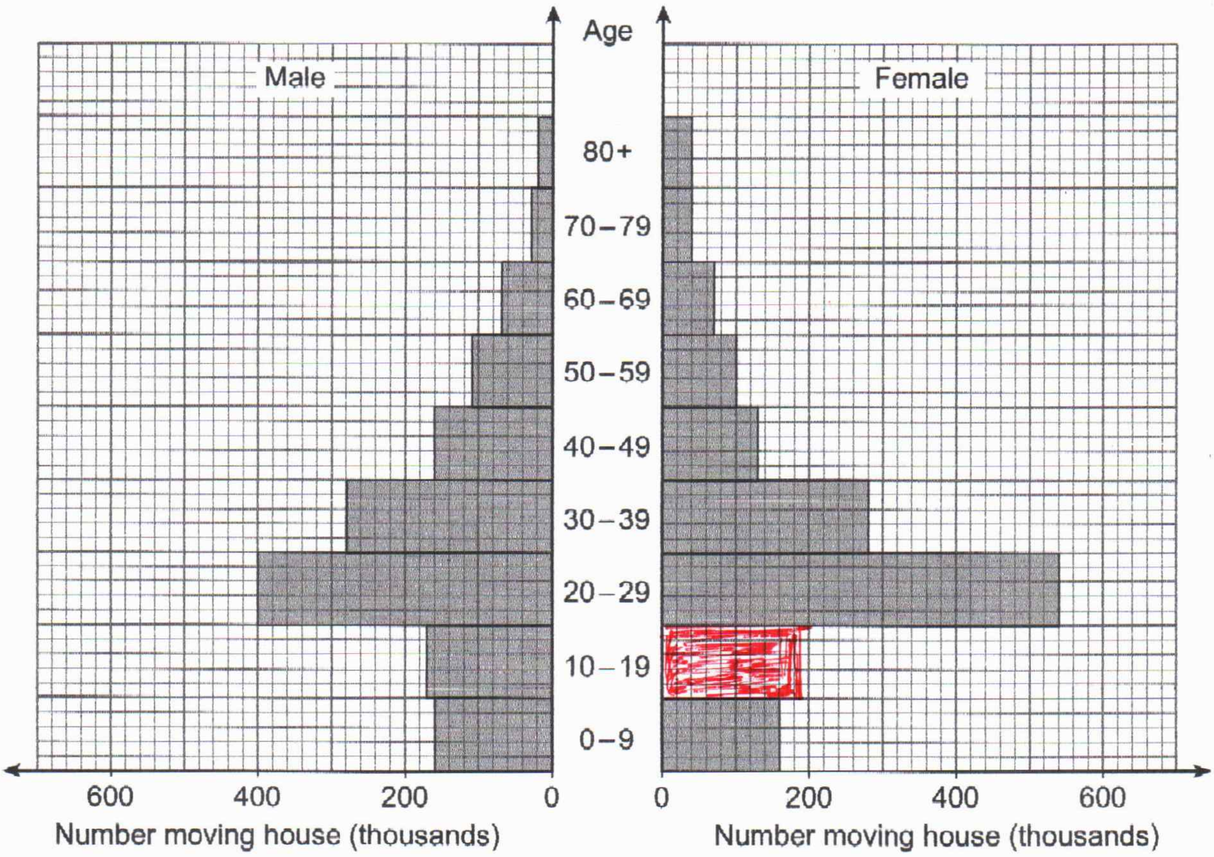
5

Turn over for the next question

Do not write outside the box

11 The population pyramid shows the ages of people who moved house in England and Wales in 2016.

One bar has not been drawn.



Total number of males moving house is 1400 thousand	Total number of females moving house is 1550 thousand

Source: ONS

11 (a) 350 000 females aged **under 20 years** moved house in 2016.

Complete the population pyramid by drawing the bar for females aged 10–19 years.

[2 marks]

Females aged (10 - 19 years) = 350,000 - 160,000

= 190,000

Do not write outside the box

11 (b) (i) Calculate the percentage of all people who moved who were aged 20–29 years.

[3 marks]

$$\left(\frac{\text{Number of males aged 20-29 years} + \text{Number of female aged 20-29 years}}{\text{Total Number of males} + \text{Total number of females}}\right) \times 100\%$$

$$= \left(\frac{400 + 540}{1400 + 1550}\right) \times 100$$

$$= \frac{940}{2950} \times 100 = 31.86$$

Answer 31.86 %

11 (b) (ii) Suggest **one** reason why such a large proportion of people moving are aged 20–29 years.

[1 mark]

Its the age where most people move to start a new job or start a family

6

Turn over for the next question

Do not write outside the box

12 The table shows some information about people with **hearing loss** in the UK.

Age	Number with hearing loss	UK population
60 years and over	8 290 000	15 590 000
Under 60 years	2 750 000	50 450 000
Total	11 040 000	66 040 000

Sources: ONS and actiononhearingloss.org.uk

12 (a) Mike says,

"The risk of hearing loss for people aged 60 years and over is about 10 times greater than the risk for people aged under 60 years."

Comment on Mike's statement.

You **must** show your working.

[3 marks]

Hearing loss for people aged 60 years and Over $= \frac{8\,290\,000}{15\,590\,000}$

$= 0.53175$

Hearing loss for People aged unders 60 years $= \frac{2\,750\,000}{50\,450\,000}$

$= 0.0545$

$\Rightarrow \frac{0.53175}{0.0545} = 9.757$

Mike's statement is correct, the risk of hearing loss for people aged 60 years and over is about 10 times greater than the risk for people aged under 60 years

12 (b) About one in nine people in the UK aged over 60 years have **sight loss**.

Calculate an estimate of the number of people in the UK aged over 60 years who have sight loss.

[1 mark]

$\frac{1}{9} \times 15,590,000$

$= 1\,732\,222$

Answer 1,732,222,

4

Do not write outside the box

13 A machine fills bottles with orange juice.

The amount of orange juice in a bottle follows a normal distribution with a mean of 500 ml and a standard deviation of 10 ml.

Range of mean = 510 - 500 = 10

13 (a) Approximately, what percentage of bottles contain **more** than 510 ml of orange juice?

Circle your answer.

[1 mark]

16%   32%   68%   84%

13 (b) The manufacturer would like **almost all** bottles to contain between 488 ml and 512 ml of orange juice.

Sophie says that this could be achieved by reducing the standard deviation to 4 ml.

Comment on Sophie's claim.

You **must** show your working.

[2 marks]

Almost all bottles gives 99.8% of the population which implies that the range of mean must be ± 3 standard deviations

Let x = standard deviation

$x = \frac{512 - 500}{3} = 4$

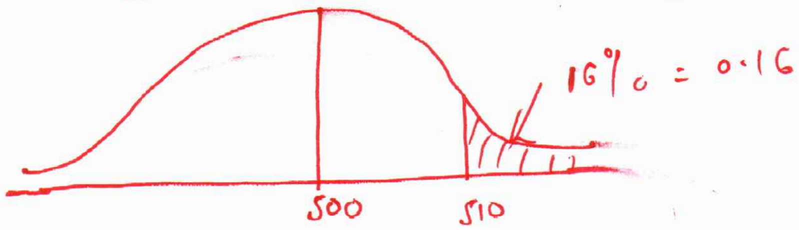
$x = \frac{500 - 488}{3} = 4$

Therefore Sophie's claim is true, the standard deviation must be reduced to 4

3

Turn over for the next question

Range of mean ± 1 standard deviation gives 68% of the population leaving 16% outside the range in each tail



Turn over ▶

Do not write outside the box

14 The table shows the value of UK imports of clothing, in £ million, from the rest of the world between 2015 Quarter 3 and 2017 Quarter 4

Some of the four-point moving averages are also shown.

Year and Quarter	Imports (£ million)	Four-point moving average
2015 Q3	4970	
2015 Q4	4730	
2016 Q1	4600	4625
2016 Q2	4200	4675
2016 Q3	5170	4725
2016 Q4	4930	4762.5
2017 Q1	4750	4870
2017 Q2	4630	4940
2017 Q3	5450	
2017 Q4	5190	

Source: ONS

14 (a) Complete the table by calculating the last moving average.

[1 mark]

$\frac{2017Q_1 + 2017Q_2 + 2017Q_3 + 2017Q_4}{4}$

$= \frac{4750 + 4630 + 5450 + 5190}{4} = 5005$ (million)

14 (b) Comment on the trend in the data.

[1 mark]

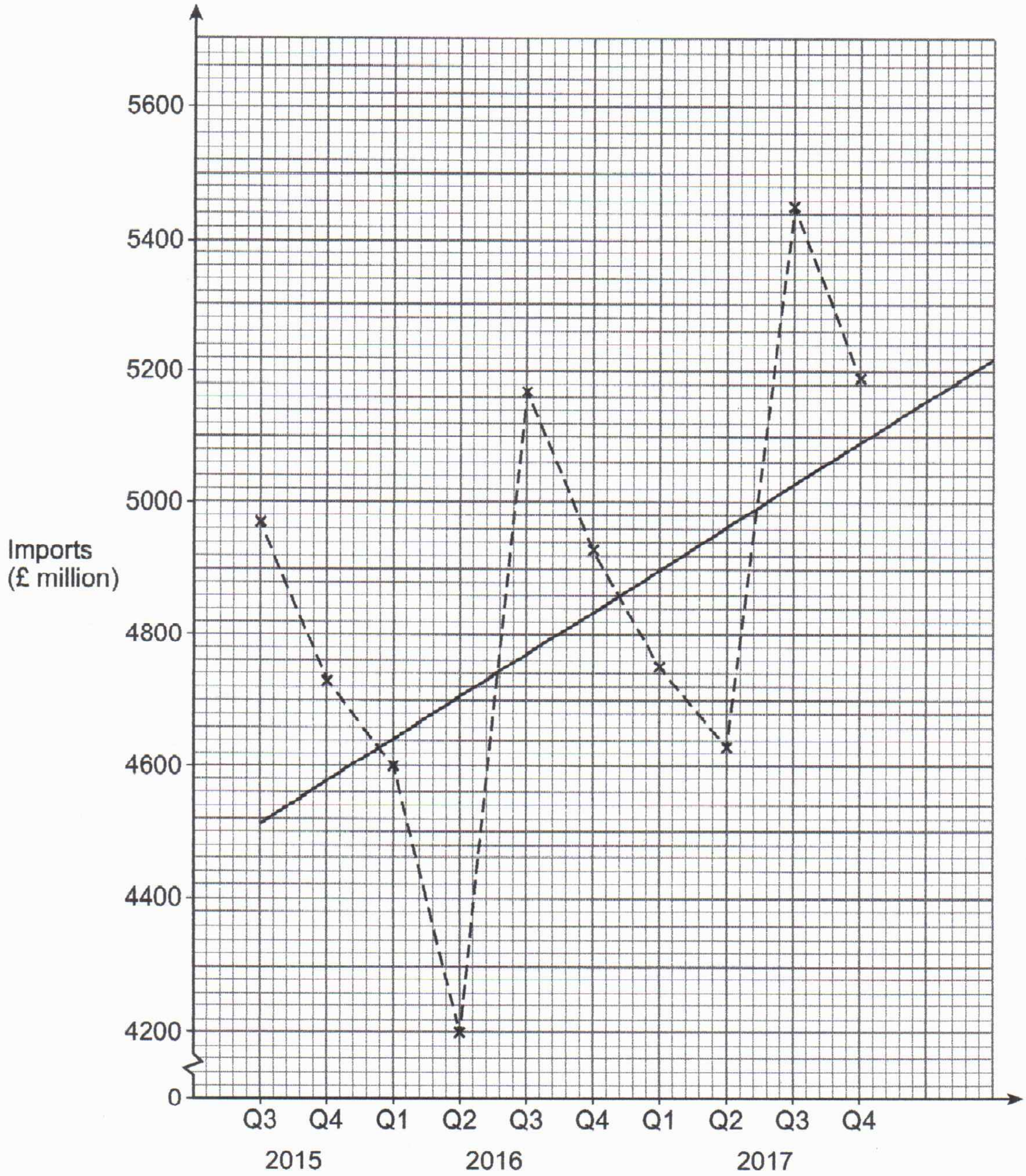
There is an increasing value of imports of clothing

Question 14 continues on the next page

Turn over ▶

The diagram shows the value of UK clothing imports in each time period.

A trend line has also been drawn.



14 (c) Make **one** comment about the seasonal variation shown in the data.

[1 mark]

- Clothing imports are lowest in Q2, that is at the start of a year
- Clothing imports are highest at Q3 which is at the end of a year.

Do not write outside the box

14 (d) The **seasonal variations** (seasonal effects) for Q1 are shown in the table.

2016 Q1	2017 Q1
−40	−150

14 (d) (i) By calculating the mean seasonal variation, predict the value of UK imports in 2018 Q1.

You **must** show your working.

[3 marks]

Estimated Mean of season variation for any season = Mean of all season variations for all seasons

Mean for 2018 Q1 = $\frac{(-40)+(-150)}{2}$ = −95

Predicted value = trend value + mean seasonal variation

= 5160 + (−95) = 5065

Answer £ 5065 million

14 (d) (ii) Write down **one** assumption that you made in making your prediction in **part (d)(i)**.

[1 mark]

Seasonal pattern remains the same

The trend continues in the same way

7

Turn over for the next question

Do not write outside the box

15 In this question you will need to use,

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

Swimmers in a competition swim two races.

Swimmers use breaststroke in Race 1 and backstroke in Race 2

The mean and standard deviation of the times in each race are shown in the table.

	Mean (seconds)	**Standard deviation (seconds)**
Race 1	45.5	2.4
Race 2	41.7	1.8

15 (a) Rachel's time in **Race 1** was 48.7 seconds.

Her standardised score in both races was the same.

Calculate Rachel's time in **Race 2**

[3 marks]

Standardised score in Race 1 $= \frac{48.7 - 45.5}{2.4}$

$= 1\frac{1}{3}$

Let Rachel's time in Race 2 $= y$

$\Rightarrow 1\frac{1}{3} = \frac{y - 41.7}{1.8}$

$y - 41.7 = 2.4$

$\Rightarrow y = 2.4 + 41.7 = 44.1$

Answer 44.1 seconds

Do not write outside the box

15 (b) Kim and Pria also swim in the competition.

Their times in each race are shown in the table below.

	Kim		Pria	
	Time (secs)	**Standardised score**	**Time (secs)**	**Standardised score**
Race 1	43.7	-0.75	44.3	-0.5
Race 2	40.5	-0.667	40.3	-0.778

Complete the table and use it to decide which race each girl swam better in.

Give a reason for each of your decisions.

[5 marks]

Kim Race 1 $= \frac{43.7 - 45.5}{2.4} = -0.75$

Kim Race 2 $= \frac{40.5 - 41.7}{1.8} = -0.667$

Pria Race 1 $= \frac{44.3 - 45.5}{2.4} = -0.5$

Pria Race 2 $= \frac{40.3 - 41.7}{1.8} = -0.778$

Kim swam better in Race 1 as her standardised score was lower in Race 1 than her score in Race 2 (-0.7

8

Turn over for the next question

⇒ Pria swam better in Race 2 as her standardised score in Race 2 was lower than her standardised score in Race 1 $(-0.778 < -0.5)$

Do not write outside the box

16 In a golf tournament, players take part in several rounds of golf.

Players try to complete the course taking as few golf strokes as possible.

Justin wants to compare the number of strokes taken by the players in the first two rounds of a tournament.

He collects data for the top 50 players.

Justin's hypothesis is,

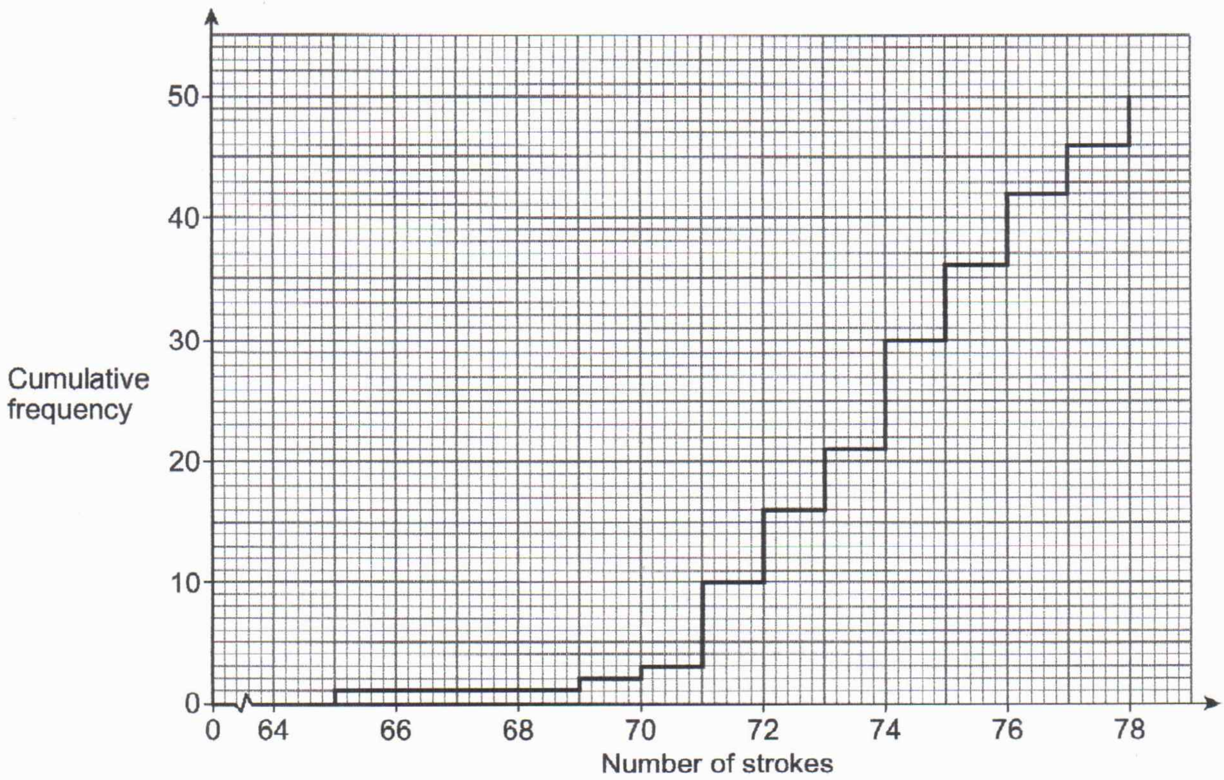
In which of the first two rounds will players take the fewer strokes on average?

16 (a) What mistake has Justin made when writing his hypothesis?

[1 mark]

A hypothesis should not be a question and Justin asked a Question

Justin draws a cumulative frequency step polygon to show the results for players in **Round 1**



16 (b) Explain why a cumulative frequency step polygon is an appropriate graph for the data. **[1 mark]**

Because the data is discrete

- Discrete data is a type of quantitative data that include figures and statistics of nondivisible single points of data that can be counted.

16 (c) Work out the percentage of players who took 72 strokes or fewer for **Round 1** **[2 marks]**

From the graph, the number of players who took 72 strokes = 16

$\frac{16}{50} \times 100\% = 32\%$

32 %

16 (d) Complete this table summarising the number of strokes taken by players in **Round 1** **[1 mark]**

Median	Lower quartile	Upper quartile
74	72	76

Upper quartile $= \frac{3}{4} \times 50 = 37.5$

From the graph the number of strokes corresponding to a cummulative frequency of $37.5 = 76$

16 (e) The lowest number of strokes taken in **Round 1** is 65

Show by calculation that this value is an outlier. **[3 marks]**

An outlier is a point of the data that lies over 1.5 Interquartile range below the lower quartile or above the upper quartile.

Outlier = Lower Quartile − 1.5 IQR

Question 16 continues on the next page

$= 72 - (76 - 72)1.5$

$= 72 - (1.5 \times 4)$

$= 72 - 6 = 66$

$65 < 66$, hence its an outlier

Turn over ▶

Do not write outside the box

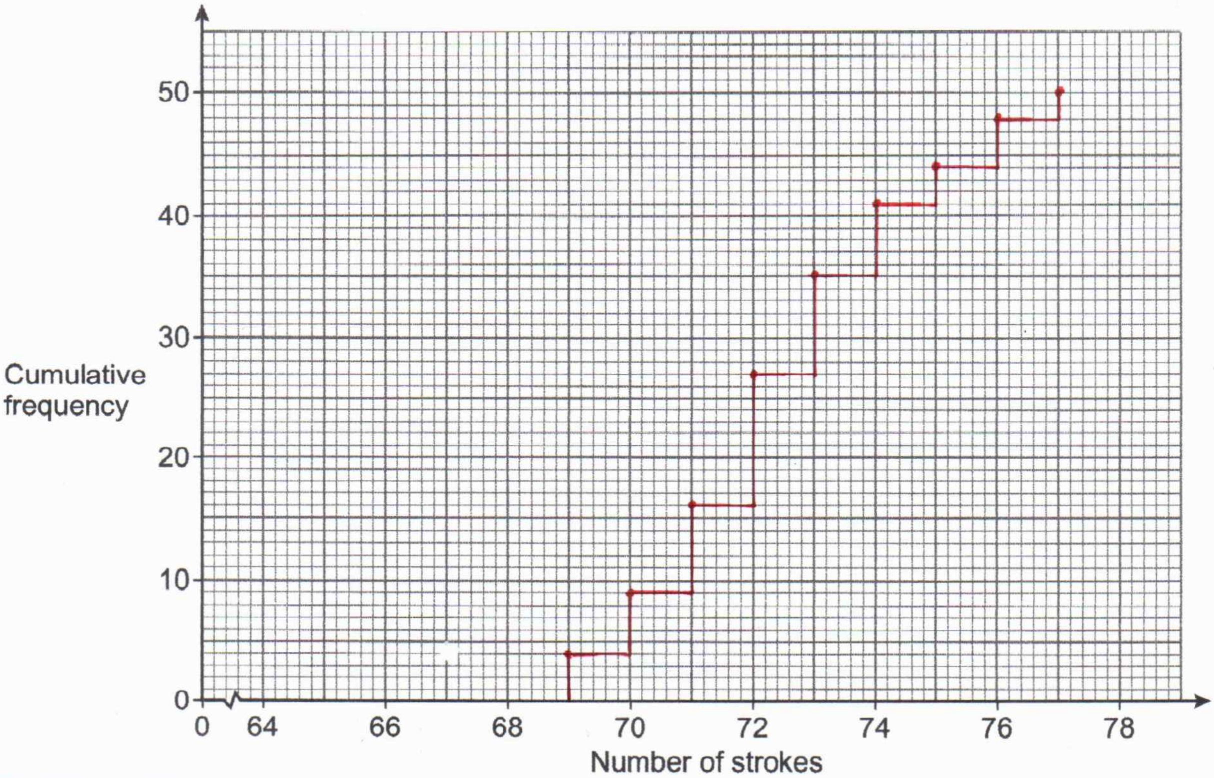
The table shows a summary of the number of strokes taken by the same players in **Round 2**

Number of strokes	Frequency
69	4
70	5
71	7
72	11
73	8
74	6
75	3
76	4
77	2

Cumulative frequency
4
9
16
27
35
41
44
48
50

16 (f) Draw a cumulative frequency step polygon to show the results for **Round 2**

[3 marks]

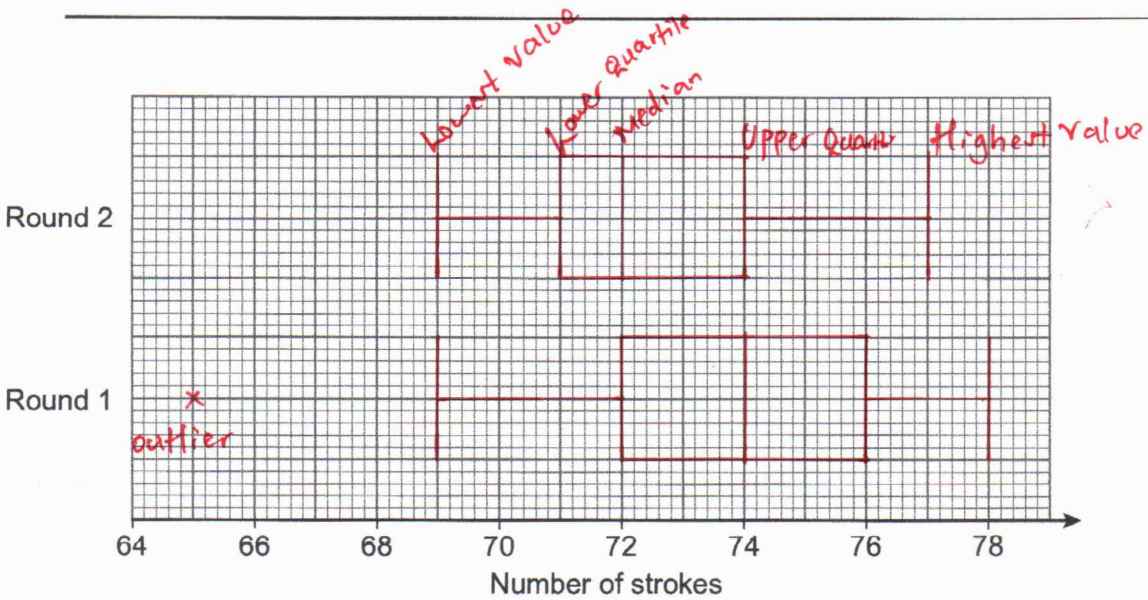


16 (g) Draw separate box plots, on the grid below, for the number of strokes in **Round 1** and **Round 2**

Mark clearly the outlier for **Round 1**

[4 marks]

65 from round 1 is an outlier.



16 (h) Compare statistically the number of strokes taken for **Round 1** and **Round 2**

[2 marks]

Median for Round 2 is equal to 72 which is less than the median value for Round 1, therefore players generally need fewer strokes on Round 2

The scores on Round 2 are less spread out because the Interquartile range for Round 2 is less than that of Round 1

16 (i) Write down a factor that could explain the difference between the number of strokes in the two rounds.

[1 mark]

- Different weather conditions when the two Rounds were played.
- Time of the day the Rounds were played

END OF QUESTIONS

18