

GCSE Statistics

Marking Guidance Summer 2024

Mel Muldowney
Christian Seagar

JustMaths



1










Foundation P1




2

F P1 Q1a/b

- 1 The incomplete pictogram gives information about the flavour and number of ice creams sold at Pradeep's cafe one Saturday morning.

Flavour	Number of ice creams
Vanilla	  
Strawberry	  
Chocolate	  

Key:  represents 8 ice creams

20 chocolate ice creams were sold on Saturday morning.

- (a) Complete the pictogram for the number of chocolate ice creams sold.

(1)

- (b) Work out the total number of ice creams sold on Saturday morning.

$$8 + 8 + 2 + 8 + 8 + 8 + 8 + 8 + 4$$

number		
1(a)	B1 For two and a half images drawn.	Accept half circle with no defined quarters. Ignore size of circles.
(b)	M1 $(2 \times 8 + \frac{1}{4} \times 8) + (3 \times 8) + 20$ or $(7 \times 8) + (\frac{1}{4} \times 8)$ A1 62	M1 for $18 + 24 + 20$ Allow one incorrect ice cream flavour total from three.

70

(2)

3

F P1 Q1c/d

- (c) B2 for e.g.
- There were more vanilla ice creams sold on the Sunday because there were 18 sold on Saturday and 45 sold on Sunday.
 - There were $(45 - 18 =)$ 27 more ice creams sold on a Sunday.









(B1 There were more (ice creams) sold on Sunday
OR
correct figures for Saturday (18) and Sunday (45) with no comparison made
OR
one correct figure for either Saturday or Sunday with a correct comparison for their values.)


The figures of 18 and 45 may be next to the pictograms.

- (d) B2 for e.g. not appropriate / no and one reason from:
- Only in the morning
 - Only on a weekend / do not know the data for the week days
 - Depends on the weather / the season / time of year / day
 - Larger sample required / 2 days is not enough
 - Large variation in data

(B1 for a correct reason with no conclusion
OR
for not appropriate with an attempt at a reason)

B2 for correct conclusion with equivalent corresponding reasoning.

Vanilla	  
Strawberry	  
Chocolate	 

Key:  represents 20 ice creams

- (c) Compare the number of vanilla ice creams sold in the cafe on Saturday morning with the number of vanilla ice creams sold in the cafe on Sunday morning.

Give a reason for your answer.

On Saturday they sold 18 vanilla ice creams
and on Sunday they sold 45 vanilla ice creams.
The pictograms look the same but the key is different

(2)

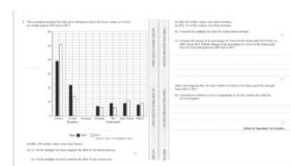
Pradeep wants to use the collected data to estimate how many ice creams of each flavour she will sell for the whole of next week.

- (d) Considering Pradeep's data decide if this is appropriate.

The data wouldn't be appropriate because
the customers could want a different flavour next week.

4

F P1 Q2c/d



<p>(c) B1B1 for each correct comparison</p> <ul style="list-style-type: none"> The percentage of visitors from France has increased (from 2003 to 2013) / France has increased (by 12%) / France has changed by (+)12% The percentage of visitors from United Kingdom has decreased (from 2003 to 2013) / UK has decreased (by 8%) / UK has changed by (-)8% 	<p>Allow for B1B1 France has increased whereas United Kingdom has decreased</p> <p>If they have stated 'increase' or 'decrease' correctly then ignore figures</p> <p>If they have stated 'changed' then figures must be correct</p> <p>Condone reference to number instead of percentage</p> <p>B0 for comparison of France v UK in 2003 or for comparison of France v UK in 2013</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

In 2003, 6% of the visitors were from Germany.

In 2013, 3% of the visitors were from Germany.

(b) Complete the multiple bar chart for visitors from Germany.

(2)

(c) Compare the change in the percentage of visitors to the theme park from France in 2003 and in 2013 with the change in the percentage of visitors to the theme park from the United Kingdom in 2003 and in 2013

France		U.K.	
2003 = 39%	% of visitors	2003 = 22%	% of visitors
2013 = 51%	increased	2013 = 14%	decreased

(2)

John is investigating how the total number of visitors to the theme park has changed from 2003 to 2013

(d) Comment on whether or not it is appropriate to use this multiple bar chart for his investigation.

it is appropriate as it visibly shows the increase or decrease in visitors.

5

F P1 Q2c/d

<p>(c) B1B1 for each correct comparison</p> <ul style="list-style-type: none"> The percentage of visitors from France has increased (from 2003 to 2013) / France has increased (by 12%) / France has changed by (+)12% The percentage of visitors from United Kingdom has decreased (from 2003 to 2013) / UK has decreased (by 8%) / UK has changed by (-)8% 	<p>Allow for B1B1 France has increased whereas United Kingdom has decreased</p> <p>If they have stated 'increase' or 'decrease' correctly then ignore figures</p> <p>If they have stated 'changed' then figures must be correct</p> <p>Condone reference to number instead of percentage</p> <p>B0 for comparison of France v UK in 2003 or for comparison of France v UK in 2013</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

In 2003, 6% of the visitors were from Germany.

In 2013, 3% of the visitors were from Germany.

(b) Complete the multiple bar chart for visitors from Germany.

(2)

(c) Compare the change in the percentage of visitors to the theme park from France in 2003 and in 2013 with the change in the percentage of visitors to the theme park from the United Kingdom in 2003 and in 2013

France has a bigger overall difference
France increased in visitors
UK decreased in visitors

(2)

<p>(d) B2 for a correct reason and conclusion e.g.</p> <ul style="list-style-type: none"> The graph only shows percentages so not appropriate / no The graph does not show the total number of visitors so not appropriate / no The graph only shows 2003 and 2013 so not appropriate / no. <p>(B1 for e.g.)</p> <ul style="list-style-type: none"> The graph only shows percentages The graph does not show the total number of visitors The graph only shows 2003 and 2013 	<p>B2 for not appropriate with a correct reason</p> <p>(B1 for a correct reason and no conclusion)</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------

John is investigating how the total number of visitors to the theme park has changed from 2003 to 2013

(d) Comment on whether or not it is appropriate to use this multiple bar chart for his investigation.

Yes because it clearly shows the difference in the amount of visitors.

(2)

6

F P1 Q2c/d

<p>(c) B1B1 for each correct comparison</p> <ul style="list-style-type: none"> The percentage of visitors from France has increased (from 2003 to 2013) / France has increased (by 12%) / France has changed by (+)12% The percentage of visitors from United Kingdom has decreased (from 2003 to 2013) / UK has decreased (by 8%) / UK has changed by (-)8% 	<p>Allow for B1B1 France has increased whereas United Kingdom has decreased</p> <p>If they have stated 'increase' or 'decrease' correctly then ignore figures</p> <p>If they have stated 'changed' then figures must be correct</p> <p>Condone reference to number instead of percentage</p> <p>B0 for comparison of France v UK in 2003 or for comparison of France v UK in 2013</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

In 2003, 6% of the visitors were from Germany.
In 2013, 3% of the visitors were from Germany.

(b) Complete the multiple bar chart for visitors from Germany.

(2)

(c) Compare the change in the percentage of visitors to the theme park from France in 2003 and in 2013 with the change in the percentage of visitors to the theme park from the United Kingdom in 2003 and in 2013

in France 2003 it was 39% and in 2013 it was 51%. In UK it was 22 in 2003 and in 2013 it was 16 percent.

<p>(d) B2 for a correct reason and conclusion e.g.</p> <ul style="list-style-type: none"> The graph only shows percentages so not appropriate / no The graph does not show the total number of visitors so not appropriate / no The graph only shows 2003 and 2013 so not appropriate / no. <p>(B1 for e.g.)</p> <ul style="list-style-type: none"> The graph only shows percentages The graph does not show the total number of visitors The graph only shows 2003 and 2013 	<p>B2 for not appropriate with a correct reason</p> <p>(B1 for a correct reason and no conclusion)</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------

John is investigating how the total number of visitors to the theme park has changed from 2003 to 2013.

(i) Comment on whether or not it is appropriate to use this multiple bar chart for his investigation.

Yes as he has all the right information on the bar chart

(2)

(2)

7

F P1 Q3b

(b) Explain why the viewing figures in the table may not be accurate.

people could be watching from websites instead of TV

(1)

<p>(b) B1 for e.g.</p> <ul style="list-style-type: none"> numbers are rounded you could not measure the number of people watching accurately viewing figures are estimates additional people watching the TV / device recorded / watched later / watched on other platform / unmonitored streams (illegal streams) / repeated episodes later TV / device left on but not watched data source is not reliable / data is secondary data 	<p>B1 for correct reason why the figures may be inaccurate</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------

4. The table shows the number of people who watched the television programme 'The Big Bang Theory' in the United Kingdom in 2011 and 2012.

Year	Number of people who watched the programme (in millions)
2011	1.2
2012	1.5

5. The table shows the number of people who watched the television programme 'The Big Bang Theory' in the United Kingdom in 2011 and 2012.

Year	Number of people who watched the programme (in millions)
2011	1.2
2012	1.5

6. Explain why the viewing figures in the table may not be accurate.

8

F P1 Q3b

(b) Explain why the viewing figures in the table may not be accurate.

they may not be accurate because they only go to 2 decimal places

(1)

<p>(b) B1 for e.g.</p> <ul style="list-style-type: none"> • numbers are rounded • you could not measure the number of people watching accurately • viewing figures are estimates • additional people watching the TV / device • recorded / watched later / watched on other platform / unmonitored streams (illegal streams) / repeated episodes later • TV / device left on but not watched • data source is not reliable / data is secondary data 	<p>B1 for correct reason why the figures may be inaccurate</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------

4. The table shows estimated data for the number of people watching the TV programme 'The Great British Bake Off' in 2017 and 2018.

Year	Total number of people watching (in millions)	Number of people watching on TV (in millions)	Number of people watching on other platforms (in millions)
2017	2.15	1.10	1.05
2018	2.30	1.20	1.10
2017	2.15	1.10	1.05
2018	2.30	1.20	1.10

Source: BBC, 2017 and 2018.

4a. What was the highest viewing figure for 'The Great British Bake Off' in 2017 and 2018?

4b. Explain why the viewing figures in the table may not be accurate.

9

F P1 Q3b

(b) Explain why the viewing figures in the table may not be accurate.

could be secondary data

(1)

<p>(b) B1 for e.g.</p> <ul style="list-style-type: none"> • numbers are rounded • you could not measure the number of people watching accurately • viewing figures are estimates • additional people watching the TV / device • recorded / watched later / watched on other platform / unmonitored streams (illegal streams) / repeated episodes later • TV / device left on but not watched • data source is not reliable / data is secondary data 	<p>B1 for correct reason why the figures may be inaccurate</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------

4. The table shows estimated data for the number of people watching the TV programme 'The Great British Bake Off' in 2017 and 2018.

Year	Total number of people watching (in millions)	Number of people watching on TV (in millions)	Number of people watching on other platforms (in millions)
2017	2.15	1.10	1.05
2018	2.30	1.20	1.10
2017	2.15	1.10	1.05
2018	2.30	1.20	1.10

Source: BBC, 2017 and 2018.

4a. What was the highest viewing figure for 'The Great British Bake Off' in 2017 and 2018?

4b. Explain why the viewing figures in the table may not be accurate.

10

F P1 Q3c

- (c) Compare the number of episodes for Emmerdale in 2016 with the number of episodes for Eastenders in 2016
Give a reason for your answer.

decreasing

(2)

(c)	<p>B2 for e.g.</p> <ul style="list-style-type: none"> There were more episodes for Emmerdale as there were 308 whereas Eastenders had 210 There were 98 more episodes for Emmerdale <p>(B1 for e.g.</p> <ul style="list-style-type: none"> There were more episodes for Emmerdale Emmerdale had 308 episodes and Eastenders had 210 There is a difference of 98) 	<p>B2 for a correct conclusion with correct comparison</p> <p>(B1 for a correct conclusion without a use of figures OR for a comparison of figures with no conclusion / incorrect conclusion)</p> <p>Figures extracted from the table must be correct</p>
-----	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

11

F P1 Q3c

- (c) Compare the number of episodes for Emmerdale in 2016 with the number of episodes for Eastenders in 2016
Give a reason for your answer.

negative correlation

(2)

(c)	<p>B2 for e.g.</p> <ul style="list-style-type: none"> There were more episodes for Emmerdale as there were 308 whereas Eastenders had 210 There were 98 more episodes for Emmerdale <p>(B1 for e.g.</p> <ul style="list-style-type: none"> There were more episodes for Emmerdale Emmerdale had 308 episodes and Eastenders had 210 There is a difference of 98) 	<p>B2 for a correct conclusion with correct comparison</p> <p>(B1 for a correct conclusion without a use of figures OR for a comparison of figures with no conclusion / incorrect conclusion)</p> <p>Figures extracted from the table must be correct</p>
-----	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

12

F P1 Q3c

- (c) Compare the number of episodes for Emmerdale in 2016 with the number of episodes for Eastenders in 2016
Give a reason for your answer.

slowly decreases from 15 to 18 from
9.87 million to 7.86 million

(2)

4. The table shows the number of episodes for Emmerdale and Eastenders in 2016.

Series	Number of episodes	Number of viewers (million)
Emmerdale	15	9.87
Eastenders	18	7.86

5. Use the information in the table to answer the questions below.

5a. How many more episodes of Emmerdale were there than Eastenders in 2016?

5b. How many more viewers watched Emmerdale than Eastenders in 2016?

5c. How many more viewers watched Emmerdale than Eastenders in 2016?

(e)

B2 for e.g.

- There were more episodes for Emmerdale as there were 308 whereas Eastenders had 210
- There were 98 more episodes for Emmerdale

(B1 for e.g.

- There were more episodes for Emmerdale
- Emmerdale had 308 episodes and Eastenders had 210
- There is a difference of 98)

B2 for a correct conclusion with correct comparison

(B1 for a correct conclusion without a use of figures OR for a comparison of figures with no conclusion / incorrect conclusion)

Figures extracted from the table must be correct

13

F P1 Q3c

- (c) Compare the number of episodes for Emmerdale in 2016 with the number of episodes for Eastenders in 2016
Give a reason for your answer.

negative correlation: the year goes up as the
viewer views down

(2)

(e)

B2 for e.g.

- There were more episodes for Emmerdale as there were 308 whereas Eastenders had 210
- There were 98 more episodes for Emmerdale

(B1 for e.g.

- There were more episodes for Emmerdale
- Emmerdale had 308 episodes and Eastenders had 210
- There is a difference of 98)

B2 for a correct conclusion with correct comparison

(B1 for a correct conclusion without a use of figures OR for a comparison of figures with no conclusion / incorrect conclusion)

Figures extracted from the table must be correct

14

F P1 Q4e

The median and range for the final 9 matches of the season are shown in the table below.

Median	Range
90	25

- (e) Use your answers to part (b) and part (d) to compare the performance of the basketball team in the first 9 matches with the performance in the final 9 matches. Give **two** comparisons and interpret **both** in context.

In the final 9 matches the team has clearly become better. This is proven as in the first 9 matches the median was 80.5 but now it is 90 showing the middle score has increased.

Also, the range has decreased from 96 to 25 showing that the team is more consistent as the best and worst score is not as far apart.

(4)

(e) B1ft B1ft B1ft B1ft for each of four comparisons or contextual interpretations

Comparison	Interpretation
e.g. • Median of first 9 games is smaller (than the median of the final 9 games)	e.g. • On average they score more points at the end of the season
• Range of the first 9 games is bigger (than the range of the final 9 games)	• They are more consistent at the end of the season

B1ft Correct comparison of medians
B1ft Correct comparison of range
B1ft One correct interpretation of median
B1ft One correct interpretation of range
fit their median in part (b) and their range in (d)
Allow equivalent/converse statements but underlined words must be seen.

15

F P1 Q4e

The median and range for the final 9 matches of the season are shown in the table below.

Median	Range
90	25

- (e) Use your answers to part (b) and part (d) to compare the performance of the basketball team in the first 9 matches with the performance in the final 9 matches. Give **two** comparisons and interpret **both** in context.

The final matches median was higher which means they have a better average score. The range is higher for the first 9 games though which means there was bigger outliers.

(4)

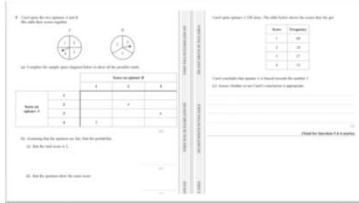
(e) B1ft B1ft B1ft B1ft for each of four comparisons or contextual interpretations

Comparison	Interpretation
e.g. • Median of first 9 games is smaller (than the median of the final 9 games)	e.g. • On average they score more points at the end of the season
• Range of the first 9 games is bigger (than the range of the final 9 games)	• They are more consistent at the end of the season

B1ft Correct comparison of medians
B1ft Correct comparison of range
B1ft One correct interpretation of median
B1ft One correct interpretation of range
fit their median in part (b) and their range in (d)
Allow equivalent/converse statements but underlined words must be seen.

16

F P1 Q5c



Carol spins spinner *A* 120 times. The table below shows the scores that she got.

Score	Frequency
1	60
2	18
3	27
4	15

Carol concludes that spinner *A* is biased towards the number 1

(c) Assess whether or not Carol's conclusion is appropriate.

Yes it is appropriate as the number 1 got much more than the rest and therefore is biased.

(c) B1 for e.g.

- There is a large number of a score of 1
- The frequency of 1 is double what it should be
- Would have expected a score of 1 to be closer to 30
- Would expect the frequencies to be similar / same

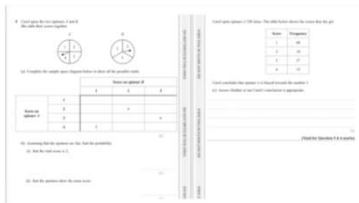
depB1 for... so Carol's conclusion is appropriate / Carol is correct

dep on B1 scored

(2)

17

F P1 Q5c



Carol spins spinner *A* 120 times. The table below shows the scores that she got.

Score	Frequency
1	60
2	18
3	27
4	15

Carol concludes that spinner *A* is biased towards the number 1

(c) Assess whether or not Carol's conclusion is appropriate.

it is a fair assessment because it shouldn't get half on '1' alone

(c) B1 for e.g.

- There is a large number of a score of 1
- The frequency of 1 is double what it should be
- Would have expected a score of 1 to be closer to 30
- Would expect the frequencies to be similar / same

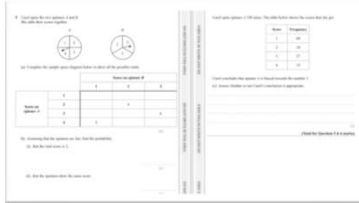
depB1 for... so Carol's conclusion is appropriate / Carol is correct

dep on B1 scored

(2)

18

F P1 Q5c



Carol spins spinner *A* 120 times. The table below shows the scores that she got.

Score	Frequency
1	60
2	18
3	27
4	15

Carol concludes that spinner *A* is biased towards the number 1

(c) Assess whether or not Carol's conclusion is appropriate.

it doesn't fit the pattern to all the other numbers

(2)

(c) B1 for e.g.

- There is a large number of a score of 1
- The frequency of 1 is double what it should be
- Would have expected a score of 1 to be closer to 30
- Would expect the frequencies to be similar / same

depB1 for... so Carol's conclusion is appropriate / Carol is correct

dep on B1 scored

19

F P1 Q6a

- 6 The manager of a gym is reviewing the current opening times of the gym. The manager thinks that if the gym is open for more hours it will affect the number of people using the gym.

(a) Suggest a hypothesis that the manager could use.

opening the gym when people are off work/school so they have time to go to gym

(1)

r		
6(a)	B1 e.g. The longer the gym stays open the more people will use it	B1 for a suitable hypothesis regarding the number of people using the gym and opening hours
		Do not accept a question

20

F P1 Q6e

The manager decides to do a pre-test of the questionnaire by giving it to a small group of people.

(e) (i) What is it called when a questionnaire is tested in this way?

fixed

(1)

(ii) Give **two** reasons why the manager might do this.

*to predict the outcome of the questionnaire and
also to change the questionnaire.*

(e)(i)	B1 Pilot	Ignore additional non-contradictory comments Accept pilot study or pilot test or pilot survey
(ii)	B1B1 for each of two from e.g. <ul style="list-style-type: none"> check response rate see if questions are understood / work or make sure the questionnaire works make sure questions get relevant answers / identify likely responses check how long it will take allows him to make improvements gain feedback on the questionnaire 	Or any other reasonable explanations Ignore additional non-contradictory comments

21

F P1 Q7a/b



(Source: www.rightmove.co.uk)

(a) Compare the proportion of different types of properties for sale in Harrogate in July 2020 with the proportion of different types of properties for sale in South Shields in July 2020

*houses had the biggest amount on both.
the second biggest amount by flats*

(2)

Adam also drew two pie charts showing the different types of properties that were for sale in July 2021 in Harrogate and in July 2021 in South Shields.

Both pie charts have the same size angle for bungalows.
Adam uses this information to reach the following conclusion.

"The numbers of bungalows for sale in Harrogate in July 2021 and in South Shields in July 2021 were the same."

(b) Assess the validity of Adam's conclusion.

*This might not be valid as he doesn't know
the population and wealth of each city.*

(2)

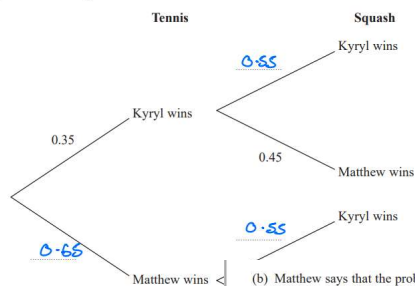
22

F P1 Q11a/b

11 Kyryl and Matthew play against each other in a game of tennis and a game of squash. In each game either Kyryl or Matthew wins.

The probability that Kyryl wins the game of tennis is 0.35
The probability that Matthew wins the game of squash is 0.45

(a) Complete the tree diagram to show this information.



(b) Matthew says that the probability of him winning both games is greater than the probability of Kyryl winning both games.
Is Matthew correct?
You must show how you get your answer.

11(a) B1 0.65 on first branch
B1 0.55, 0.55, 0.45 on second branches

Accept fractions, decimals and percentages.

(b) M1 0.35×0.55 (= 0.1925) OR ' 0.65×0.45 ' (= 0.2925)
A1ft 0.19(25)
A1ft 0.29(25)

M1 for product of their probabilities from their tree.

A1ft 0.19(25) or follow through their tree
A1ft 0.29(25) or follow through their tree

depB1ft Matthew is correct (as $0.2925 > 0.1925$)

dep on M1 and two probabilities calculated.
Their probabilities must be $0 < p < 1$

Matthew is correct there is a much higher probability of Matthew winning the first game 65% chance as he has got more of a chance than Kyryl of winning both games (4)

(Total for Question 11 is 6 marks)

23

Foundation / Higher
Crossover
P1



24

F P1 Q8 bii / iii and c

(ii)	B1 for any one from <ul style="list-style-type: none"> Data is grouped / is in intervals Use midpoint values True values are not known 	Accept equivalent wording but reference to rounding or it is an estimate is B0
(iii)	B1 for any one from <ul style="list-style-type: none"> splitting the data into more groups / smaller widths for the groups Use raw / ungrouped data Check the accuracy of measurements / use another source 	Accept increase sample size / find more lengths. B0 for collect primary data.
(c)	B2 for appropriate and a correct reason <ul style="list-style-type: none"> (continuous) grouped data can show the distribution of lengths (B1 for e.g. grouped data with no / incorrect conclusion) OR B2 for not appropriate and a correct reason <ul style="list-style-type: none"> poor class widths variable frequencies / a lot more lengths in $0 \leq l < 50$ (B1 for poor class widths / large variation in frequencies with no / incorrect conclusion)	B2 for assessing the appropriateness of using a frequency polygon with consistent reason (B1 for a correct reason and no conclusion / incorrect conclusion) Ignore additional non-contradictory statements. Ignore reference to alternative diagrams.

(ii) Explain why your answer to part (b)(i) is only an estimate.

as the data used to form this is from the internet
put into a grouped freq table

(1)

(iii) How could Emily have improved the accuracy of her answer to part (b)(i)?

repeat many hundreds calculate means

(1)

Emily plans to use a frequency polygon to represent the lengths of the fjords.

(c) Discuss whether or not a frequency polygon would be an appropriate diagram to use.

Yes it could show patterns and trends

(2)

25

F P1 Q8 b ii / iii and c

(ii)	B1 for any one from <ul style="list-style-type: none"> Data is grouped / is in intervals Use midpoint values True values are not known 	Accept equivalent wording but reference to rounding or it is an estimate is B0
(iii)	B1 for any one from <ul style="list-style-type: none"> splitting the data into more groups / smaller widths for the groups Use raw / ungrouped data Check the accuracy of measurements / use another source 	Accept increase sample size / find more lengths. B0 for collect primary data.
(c)	B2 for appropriate and a correct reason <ul style="list-style-type: none"> (continuous) grouped data can show the distribution of lengths (B1 for e.g. grouped data with no / incorrect conclusion) OR B2 for not appropriate and a correct reason <ul style="list-style-type: none"> poor class widths variable frequencies / a lot more lengths in $0 \leq l < 50$ (B1 for poor class widths / large variation in frequencies with no / incorrect conclusion)	B2 for assessing the appropriateness of using a frequency polygon with consistent reason (B1 for a correct reason and no conclusion / incorrect conclusion) Ignore additional non-contradictory statements. Ignore reference to alternative diagrams.

(ii) Explain why your answer to part (b)(i) is only an estimate.

the table shows class width not exact values

(1)

(iii) How could Emily have improved the accuracy of her answer to part (b)(i)?

used all of the exact values of each fjord
to work out the mean

(1)

Emily plans to use a frequency polygon to represent the lengths of the fjords.

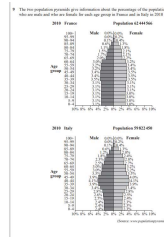
(c) Discuss whether or not a frequency polygon would be an appropriate diagram to use.

No because it can't represent the data properly

(2)

26

F P1 Q9



Tommy is investigating how the populations of Italy and France differ in 2010 and 2019 using the two population pyramids to reach the following two conclusions.

Conclusion 1: The percentage of people aged 50-54 was lower in France than the percentage of people aged 50-54 in Italy.

Conclusion 2: The number of males aged 40-44 in France was greater than the number of males aged 40-44 in Italy.

Tommy's two conclusions could show clearly the value of any statistics you use in your answer.

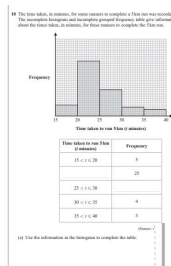
France	Italy
50-54 3.2%	50-54 3.8%
40-44 3.4%	40-44 4.1%
3.4% of 62,444,566	4.1% of 59,822,450
= 212,315	= 2,452,720

The percentages of people aged 50-54 in France was lower than the percentage of people aged 50-54 in Italy as France had 3.2% whilst Italy had 3.8%. Tommy's first conclusion was right, however for the number of males aged 40-44 in France was not greater than the number of males in Italy as Italy had 2,452,720 more males aged 40-44 than France.

(Total for Question 9 is 5 marks)

27

F P1 Q10c



(c) Identify and interpret the skew shown on the histogram.

it has a positive correlation.

(2)

(c) B1 for identifying positive skew

B1 for interpretation e.g.

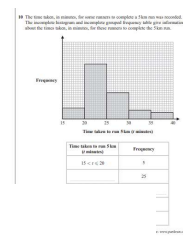
- times greater than the median are more spread out (than times less than the median)
- more than half of the runners take less than the mean time to run 5km
- the mean time is greater than the median time
- the times are mainly at the lower end of the distribution

B1 for identifying the correct skew
B0 for positive correlation

B1 for correct interpretation of skew
Accept interpretation not in context e.g. the mean is greater than the median

28

F P1 Q10c



(c) Identify and interpret the skew shown on the histogram.

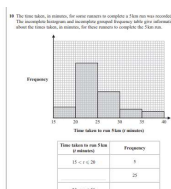
it is a positive skew as more people took
20 - 25 minutes

(2)

<p>(c) A1 20</p> <p>B1 for identifying positive skew</p> <p>B1 for interpretation e.g.</p> <ul style="list-style-type: none"> times greater than the <u>median</u> are more spread out (than times less than the median) more than half of the runners take less than the <u>mean</u> time to run 5km the <u>mean</u> time is greater than the <u>median</u> time the times are mainly at the lower end of the distribution 	<p>B1 for identifying the correct skew B0 for positive correlation</p> <p>B1 for correct interpretation of skew Accept interpretation not in context e.g. the <u>mean</u> is greater than the <u>median</u></p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

29

F P1 Q10c



(c) Identify and interpret the skew shown on the histogram.

There is a positive skew as the median is more
to the left than the right of the diagram.

(2)

<p>(c) A1 20</p> <p>B1 for identifying positive skew</p> <p>B1 for interpretation e.g.</p> <ul style="list-style-type: none"> times greater than the <u>median</u> are more spread out (than times less than the median) more than half of the runners take less than the <u>mean</u> time to run 5km the <u>mean</u> time is greater than the <u>median</u> time the times are mainly at the lower end of the distribution 	<p>B1 for identifying the correct skew B0 for positive correlation</p> <p>B1 for correct interpretation of skew Accept interpretation not in context e.g. the <u>mean</u> is greater than the <u>median</u></p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

30

Higher P1



31

H P1 Q1b

You must write down all the steps in your working.

1. The table gives the total labour force and the unemployment rate for the UK in 2017 and in 2018.

Year	Total labour force (million)	Unemployment rate (%)
2017	33.88	4.38
2018	32.43	4.89
2019	34.84	

(Source: HM Treasury, Office for National Statistics)

In 2019 the total number of unemployed people was 1.29 million.

Unemployment rate = $\frac{\text{Number of unemployed people}}{\text{Total labour force}}$

(a) (i) Using the formula above, work out the unemployment rate in 2017. Give your answer correct to 2 decimal places and write your answer in the box.

(ii) Using your answer to part (a)(i), what conclusion can be drawn about the unemployment rate in the UK between 2017 and 2018?

Bob says, without doing any calculations, that the total number of people unemployed decreased from 2017 to 2018.

(b) Using the data in the table, assess Bob's claim.

Bob's claim may not be accurate as the total labour force decreased as well as the unemployment rate.

Bob says, without doing any calculations, that the total number of people unemployed decreased from 2017 to 2018

(b) Using the data in the table, assess Bob's claim.

Bob's claim may not be accurate as the total labour force decreased as well as the unemployment rate

(2)

(Total for Question 1 is 5 marks)

(b)	<p>B2 for a correct decision and complete reason e.g.</p> <ul style="list-style-type: none"> Bob is correct as the total workforce and unemployment rate is lower in 2018 compared to 2017 <p>(B1 for e.g.</p> <ul style="list-style-type: none"> Bob is correct as the total workforce is lower in 2018 compared to 2017 Bob is correct as the unemployment rate is lower in 2018 compared to 2017 the total workforce and unemployment rate is lower in 2018 with no conclusion or incorrect conclusion Bob is correct with 1.48(044) million and 1.296(4) million) 	<p>B2 for a complete assessment of the claim together with reason</p> <p>(B1 for a complete reason and no or incorrect conclusion or for a correct decision with partial reason).</p> <p>Not 1.29 million (2019 figure)</p>
-----	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

32

H P1 Q5 d/e

5 The management of a factory is considering changing the working hours of their employees. Muhammad and Rose want to get the views of the employees in the factory. Employees in the factory work on the production line or in the warehouse or in the office.

20 employees work on the production line.
15 employees work in the warehouse.
25 employees work in the office.

Muhammad plans to use a questionnaire. He plans to take a sample of the employees and ask them the questions on his questionnaire. For his sample, he decides to ask all of the employees who work on the production line.

(a) (i) Name this sampling technique. (1)

(ii) Give two reasons why using this sampling technique may not be appropriate. (2)

1 _____

2 _____

Muhammad wants to find out how many extra hours each employee would be willing to work each week.

(b) Design a closed question that Muhammad could use in his questionnaire. (2)

Rose decides to take a 10% systematic sample of all the 60 employees in the factory.

(d) Describe in detail how this sample could be selected.

Take each employee and assign them a number 0 - 60. Then using a random number generator choose 10% so 6 numbers and then you have your sample

Rose plans to use a face-to-face interview.

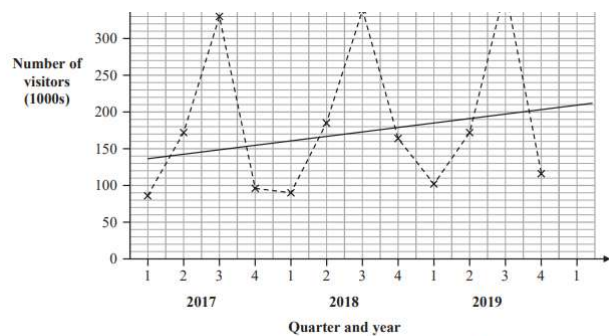
(e) How would using a face-to-face interview rather than a questionnaire improve the quality of the responses?

you are being told it instead of reading it on a piece of paper.

(d)	B1 get a list of all the 60 employees / get a sample frame / number all the employees B1 sample every 10 th person (can be implied by e.g. 5, 15, 25) B1 use a random starting point between 1 and 10. OR B1 get a list of employees / get a sample frame B1 use a spreadsheet to randomise the order of names B1 pick every 10th person	Allow for random starting point between 1 and a number greater than 10 / no upper limit.
(e)	B1 for e.g. • questions can be explained • can ask follow-up questions / get more detailed answers / clarify answers	B0 for reference to increased number of responses. B0 for reference to feeling pressured / less likely to lie. Do not accept e.g. 'open responses can be used'.

33

H P1 Q6a



A trend line has been drawn on the time series graph.

(a) Describe and interpret the trend shown by the graph.

There is a positive / rising trend meaning that the number of visitors is increasing

umber	Answer	Additional guidance
6(a)	B2 for e.g. the number of visitors (to Canada) is increasing (over time). (B1 for increasing/upwards/positive/correct description of the trend with missing or incorrect contextual interpretation)	B2 for a correct description of the trend with contextual interpretation (B1 for increasing/correct description of the trend with missing or incorrect contextual interpretation) Allow rising. Do not allow positive correlation alone , but condone if accompanied by e.g. increasing / upwards trend. Ignore reference to figures.

34

H P1 Q6c

- (c) (i) Work out the mean seasonal variation for Quarter 1.
Give your answer correct to 1 decimal place.

$$135 - 85 = 50$$

$$190 - 90 = 100$$

$$185 - 100 = 85$$

$$\frac{50 + 100 + 85}{3}$$

$$68.3333$$

68.3

thousand
(2)

(c)(i)	M1 $\frac{+(-50 + -70 + -85)}{3}$ A1 -68.3	Working may be seen on graph. Allow ± 5 on each reading. Allow answers in the range -65 to -70 Working may be in thousands.
	B1ft e.g. on average quarter 1 has '68.3' (thousand) fewer	B1 for a correct interpretation in context .

35

H P1 Q7a

- 7 Roberta is investigating how the ages of brides getting married in the UK has changed from 2003 to 2013.
She collects official data from the internet using the website 'Office for National Statistics'.

- (a) Explain why this website will give reliable data.

it is a respected company

(1)

7(a)	B1 for e.g. <ul style="list-style-type: none"> ONS is known to have quality assurance standards / data is checked trustworthy source collects large amounts of data 	Accept e.g. government data. Accept e.g. data for everyone, data for whole country (implies large amounts of data) Do not accept e.g. reliable source (reliability asked in question), official data, national website on its own.
------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

36

H P1 Q7c

Roberta wants to compare the proportion of brides in each age group and the total number of brides in each age group by presenting the results in charts for 2003 and 2013. She discusses how to present the results with Andria.

Andria thinks that they should use pie charts.

Roberta thinks that they should use comparative pie charts.

(c) What advice would you give to Andria and to Roberta on their choice of charts?

Comparative pie charts also shows the difference in sizes and not just in proportion so it will show there was more people in 2003

- (c) B1 pie charts appropriate as want to compare proportions (of brides in different age groups) / pie charts allow us to compare proportions (of brides in different age groups)
- B2 Comparative pie charts more appropriate / better or Roberta correct plus a reason e.g.
- as totals are different
 - more brides in 2003
- (B1 for e.g. totals are different / more brides in 2003 with no conclusion or incorrect conclusion)
- OR
- B2 Comparative pie charts are not appropriate / not necessary or Andria plus reason
- totals are similar
- (B1 for e.g. totals are similar with no conclusion or incorrect conclusion)

B1 for indicating a pie chart is appropriate together with a correct reason. This may be as part of a comment on comparative pie charts.

B2 for a decision on the appropriateness of comparative pie charts with a correct reason (B1 for a correct reason with no conclusion or incorrect conclusion).

Note: indication that comparative pie charts are more appropriate may be e.g. 'Andria should use comparative pie charts' (as Andria originally planned only to use pie charts).

(3)

37

H P1 Q10

The incomplete table below gives their times and standardised scores.

	Dominic		Kai	
	Time (seconds)	Standardised score	Time (seconds)	Standardised score
100 m race	13.35	-0.5	13.58	-0.4
400 m race	56.12	-0.2	58.82	0.3

Dominic and Kai make the following conclusions.

- Dominic concludes that he performed better in the 400 m race compared to the 100 m race.
- Kai concludes that he finished over one second slower than Dominic in the 400 m race.

Complete the table and assess Dominic's and Kai's conclusions. Give a reason for each of your decisions.

Dominic is incorrect as his standardised score per 100m is lower than his standardised score per 400m so he performed better in 100m compared to everyone else.
Kai is correct as he finished 2.7 seconds slower than Dominic in 400m

(Total for Question 10 is 6 marks)

38

Foundation P2



39

F P1 Q2c

Jenny thinks that there are a lot of people in her fitness group who are exercising less than 2 days per week as there is a total of 10 people who used the gym on 0 days or 1 day per week.

(c) Explain why Jenny might **not** be correct.

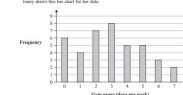
Because there are 30 more people training more than 2 days a week.

(1)

<p>(c) B1 e.g.</p> <ul style="list-style-type: none"> 30 people are exercising 2 or more days a week 10 people is (only) 25% of her fitness group / 10 out of 40 is not a lot This only shows exercise at the gym, they might exercise elsewhere more than 10 people used the gym for 2 or more days There are 40 people, only 10 do exercise on 0 or 1 day 	<p>(1)</p> <p>A comparison of the those using the gym less than 2 days and those using 2 days or more is sufficient. 10 and 30 need not be seen.</p> <p>Allow 'the majority of people exercise more than two times a week'</p> <p>Condone sample of 40 is too small</p> <p>Do not allow</p> <ul style="list-style-type: none"> Gym use varies week by week The question may have been misunderstood 10 people is not a lot
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

2. Jenny is investigating how many days per week people use a gym.

She asks the 40 people in her fitness group how often they use the gym each week. Jenny thinks there are a lot of people who use the gym 0 or 1 day per week.



One of these people is chosen at random.

(a) Find the probability that the person uses the gym exactly 2 days per week.

(b) What is the modal number of days to use the gym each week?

Jenny thinks that there are a lot of people in her fitness group who are exercising less than 2 days per week as there is a total of 10 people who used the gym on 0 days or 1 day per week.

(c) Explain why Jenny might not be correct.

40

F P1 Q2c

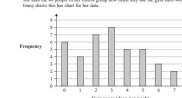
Jenny thinks that there are a lot of people in her fitness group who are exercising less than 2 days per week as there is a total of 10 people who used the gym on 0 days or 1 day per week.

(c) Explain why Jenny might **not** be correct.

There is a higher frequency of people using the gym more than 3 hours a week compared to 2 or less (1)

<p>(c) B1 e.g.</p> <ul style="list-style-type: none"> 30 people are exercising 2 or more days a week 10 people is (only) 25% of her fitness group / 10 out of 40 is not a lot This only shows exercise at the gym, they might exercise elsewhere more than 10 people used the gym for 2 or more days There are 40 people, only 10 do exercise on 0 or 1 day 	<p>A comparison of the those using the gym less than 2 days and those using 2 days or more is sufficient. 10 and 30 need not been seen.</p> <p>Allow 'the majority of people exercise more than two times a week'</p> <p>Condone sample of 40 is too small</p> <p>Do not allow</p> <ul style="list-style-type: none"> Gym use varies week by week The question may have been misunderstood 10 people is not a lot
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

2 Jenny is investigating how many days per week people use a gym. She asks the 40 people in her fitness group how often they use the gym each week. Jenny thinks that there are a lot of people who use the gym 0 or 1 day per week.



One of these people is chosen at random.

(a) Find the probability that the person uses the gym exactly 2 days per week.

(b) What is the modal number of days to use the gym each week?

Jenny thinks that there are a lot of people in her fitness group who are exercising less than 2 days per week as there is a total of 10 people who use the gym on 0 days or 1 day per week.

(c) Explain why Jenny might **not** be correct.

41

F P1 Q3e

Ben wants to use an average to summarise the data.

(e) Which of the mode or the median would be more appropriate? Give a reason for your answer.

Median is the middle value so it's an average. (2)

<p>(e) B1 for e.g. median as there is more than one mode (B1 for e.g. median with an attempt at a reason OR there is more than one mode)</p>	<p>B2 ft for identifying the appropriate average together with a reason (B1 for identifying median with an attempt at a reason OR for a reason without a decision)</p> <p>B0 for median with no attempt at reason</p>
----------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

42

F P1 Q5

<p>B1 Comments on data collection:</p> <ul style="list-style-type: none"> Taking a random sample would mean each test centre in a region had an equal chance/unbiased chance of being selected Visiting the test centres would take too long / cost too much / isn't practical It would be quicker / easier / cheaper to collect data from the internet / secondary data May not be able to get the information required by asking at the test centres A good idea to ask for data for the same month for each test centre Should collect data in more than one month Should include information on sample size Good to include all of the regions/representative of the regions <p>B1ft dep for appropriate comment on appropriateness on method of data collection consistent with the observations made.</p> <p>B1 Comments on calculations:</p> <ul style="list-style-type: none"> Use of an average is a good way to represent the waiting time overall for each region. Claire should specify which average she plans to use. Calculating the mean/median of each region. Mode would not be a suitable average to use. The range would give an idea of the spread of waiting times within each region. <p>B1ft dep for appropriate comment on appropriateness on calculations consistent with the observations made.</p>	<p>B1 for a correct comment relating to the methods of data collection</p> <p>B1ft for comment on appropriateness of data collection consistent with their observations Dependent on previous B mark being awarded for data collection</p> <p>B1 for a correct comment relating to the calculations</p> <p>B1ft for comment on appropriateness of calculations consistent with their observations Dependent on previous B mark being awarded for calculations</p>	(6)
<p>B1 Comments on diagrams:</p> <ul style="list-style-type: none"> A bar chart would make it easier to <u>compare</u> the average waiting times for the different areas. A bar chart is not suitable for time as it is continuous data. A histogram or frequency polygon would be better to show continuous data. A pie chart would not be a suitable way represent the type of data for the range of waiting times for the different areas. <p>B1ft dep for appropriate comment on appropriateness on diagrams consistent with the observations made.</p>	<p>B1 for a correct comment relating to the diagrams</p> <p>B1ft for comment on appropriateness of diagrams consistent with their observations Dependent on previous B mark being awarded for diagrams</p>	

- 5 Claire is planning an investigation into the length of time that a learner has to wait for a driving test.

She wants to find out about how waiting time varies in different regions of the UK.

Here is her plan for data collection, for calculations and for diagrams.

Data collection

Visit a random sample of driving test centres in each region to ask for their waiting time in June.

Calculations

Calculate the average waiting time for each region for June.
Calculate the range of the waiting times for each region for June.

Diagrams

Draw a bar chart showing the average waiting time for each region in June.
Draw a pie chart showing the range of waiting times for each region in June.

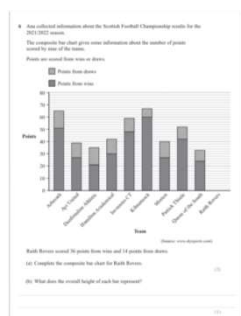
Discuss whether Claire's plans for data collection, for calculations and for diagrams are appropriate.

They are appropriate however this would take too long. Use secondary data which will be quicker and easier.

She can also calculate a mean for each region and add it together

45

F P1 Q6b



- (b) What does the overall height of each bar represent?

The amount of wins and draws.

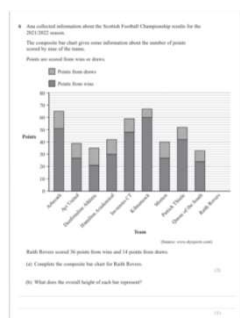
(1)

B1 e.g. the overall/total points scored/points from wins and draws

(1)

46

F P1 Q6b



(b) What does the overall height of each bar represent?

amount of goals scored

(1)

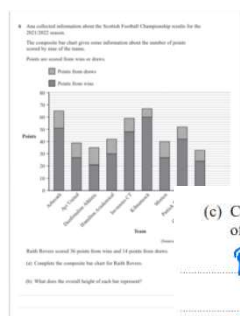
(1)

B1 e.g. the overall/total points scored/points from wins and draws

(1)

47

F P1 Q6c



(c) Compare the points scored by Dunfermline Athletic with the points scored by Queen of the South.

Dunfermline scored 2 more points in total than Queen of the South.

Queen of the South scored 3 more points from wins and 5 less points in draws

B1 e.g. Queen of the South scored more points for winning than Dunfermline Athletic

B1 e.g. Dunfermline Athletic scored more points for drawing than Queen of the South

B1 e.g. Dunfermline Athletic scored more points overall than Queen of the South

B1 for a correct **comparison** of points for winning
 B1 for a correct **comparison** of points for drawing
 B1 for a correct **comparison** of total points scored

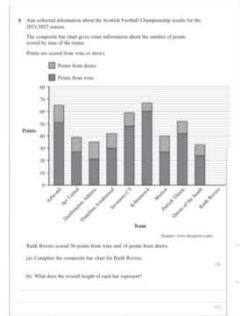
Ignore any numbers in written response.

(3)

(3)

48

F P1 Q6c



(c) Compare the points scored by Dunfermline Athletic with the points scored by Queen of the South.

QOTS scored 24 points from wins whereas DA scored around 21 points from wins.

DA scored 12 points from draws whilst QOTS scored 9 points from draws

(3)

B1 e.g. Queen of the South scored more points for winning than Dunfermline Athletic

B1 e.g. Dunfermline Athletic scored more points for drawing than Queen of the South

B1 e.g. Dunfermline Athletic scored more points overall than Queen of the South

B1 for a correct **comparison** of points for winning

B1 for a correct **comparison** of points for drawing

B1 for a correct **comparison** of total points scored

Ignore any numbers in written response.

(3)

49

F P1 Q7 a/b/c

7 Chris is a manager at a theme park.

He wants to find out what food options visitors would like to be able to buy in the theme park.

(a) State the population for this investigation.

visitors in the park.

(1)

Chris decides that he will take a convenience sample of visitors in the section of the park selling food.

(b) Explain what is meant by a convenience sample.

something easy and without much thought such as standing at the theme park entrance and asking the first 100 people.

(1)

(c) Give one disadvantage of using a convenience sample.

The sample is unlikely to be representative of the population

(1)

7(a) B1 for all the visitors (to the theme park)

Must include reference to all.

(b) B1 e.g.

- sampling the people who are available at the time
- sampling only those who stop to answer your questions
- sampling those only closest to you
- sampling the first people you see
- sampling people easy to access

B0 for 'all people buying food'
B1 for a definition of opportunity sampling

Allow a description of how a convenience sample could take place
E.g. 'sampling people queuing up in a line'/'sat down eating'

(c) B1 e.g.

- Not representative
- (May be) biased

B1 for a disadvantage of a convenience sample

(1)

50

F P1 Q7 d/e

(d)	<p>B2 for two comments from</p> <ul style="list-style-type: none"> (a data collection sheet makes it) easy to analyse responses / put in graphs/ can identify the most liked product There are too few options (e.g. no 'burgers') Visitors may choose more than one option Visitors may not buy/like food at the theme park Chinese and curry are vague options Other should be included <p>(B1 for one comment from the list)</p>	<p>B2 for two comments on the appropriateness of using this data collection sheet. (B1 for one comment on the appropriateness of using this data collection sheet)</p>	(2)
(e)	<p>B2 for 'not suitable' as data is qualitative (not numerical / quantitative) (B1 for not suitable with an attempt at a reason OR for identifying that data is qualitative)</p>	<p>B2 for a complete answer assessing that a stem and leaf diagram is not suitable with a correct reason (B1 for not suitable with an attempt at a reason OR for identifying that data is qualitative)</p>	(2)

(d) Discuss whether this data collection sheet is appropriate.

You should consider how Chris might use the data and describe any problems he might have when he uses the data collection sheet.

it shows a clear result after the data is collected but does not leave room for added opinions

(2)

Chris suggests using a stem and leaf diagram to represent the data that he collects.

(e) Discuss whether or not this would be a suitable diagram to represent his data.

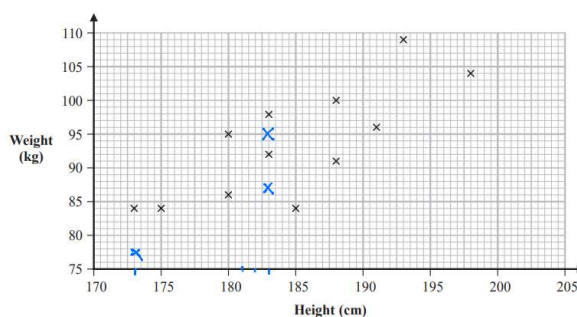
it is not appropriate because its not the type of diagram to use for the data hes collecting

(2)

51

F P1 Q8b

the Wales rugby squad.



Player	A	B	C
Height (cm)	183	183	173
Weight (kg)	95	87	78

(b) Complete the scatter diagram by plotting the points for players A, B and C.

B2 for all three points plotted correctly
(B1 for one or two points plotted correctly)

accept equivalent

(2)

B1 for positive correlation

B1 for identifying the type of

(2)

52

F P1 Q8g

Timur uses the information in the table to conclude that the weight of the England rugby squad Backs increases faster than the weight of the Wales rugby squad Backs as their height increases.

(g) Assess the validity of Timur's conclusion with reference to the statistical results.

This is a valid assessment to make as the gradient for England is steeper than for Wales

(2)

<p>B2 for e.g.</p> <ul style="list-style-type: none"> Timur <u>is correct</u> as the gradient of the line of best fit of the England rugby squad backs is greater/steeper Timur is <u>not correct</u> as the gradient of the lines of best fit is <u>similar</u> <p>(B1 for e.g.</p> <ul style="list-style-type: none"> the gradient of the line of best fit of the England rugby squad backs is greater the gradient of the lines of best fit is similar for each extra centimetre of height the weight of the Wales rugby backs increases by 0.96kg and the weight of the England rugby backs increases by 1.02kg 	<p>B2 for assessment of the validity of the conclusion with supporting reason (B1 for comparison of the gradients of the lines of best fit OR contextual interpretation with no or incorrect assessment of validity)</p>	<p>(2)</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------

F P1 Q8g

Timur uses the information in the table to conclude that the weight of the England rugby squad Backs increases faster than the weight of the Wales rugby squad Backs as their height increases.

(g) Assess the validity of Timur's conclusion with reference to the statistical results.

He is correct as England has a higher gradient of line of best fit meaning the weight increases faster as height increases

(2)

<p>B2 for e.g.</p> <ul style="list-style-type: none"> Timur <u>is correct</u> as the gradient of the line of best fit of the England rugby squad backs is greater/steeper Timur is <u>not correct</u> as the gradient of the lines of best fit is <u>similar</u> <p>(B1 for e.g.</p> <ul style="list-style-type: none"> the gradient of the line of best fit of the England rugby squad backs is greater the gradient of the lines of best fit is similar for each extra centimetre of height the weight of the Wales rugby backs increases by 0.96kg and the weight of the England rugby backs increases by 1.02kg 	<p>B2 for assessment of the validity of the conclusion with supporting reason (B1 for comparison of the gradients of the lines of best fit OR contextual interpretation with no or incorrect assessment of validity)</p>	<p>(2)</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------

F P1 Q9b

9 Mobeen is investigating whether there is a difference in the amount of time spent reading by pupils in Green Park school and pupils at Golden Plains school. He uses a census of all of the pupils at each school. Each pupil is asked to record the amount of time spent reading in a week. Mobeen then collects this information from each student through an online database. Part of the database is shown below.

	School	Time spent reading
1	Green Park	3 hours and 10 minutes
2	Golden	2.8 hours
3	GP	40
4	GREEN PARK	3630
5	Golden Plains	$3\frac{1}{2}$ h
6	Green park	About 3 hours
7	Green park school	None
8	-	80
9	Golden plains	1.3h

(a) Give two reasons why the data should be cleaned before processing.

Mobeen wants to compare the data for Green Park school with the data for Golden Plains school. Once the data has been cleaned Mobeen plans to use all of the times to draw a single box plot.

(b) Explain why this is **not** an appropriate thing to do.

Mobeen wants to compare the data for Green Park school with the data for Golden Plains school.

Once the data has been cleaned Mobeen plans to use all of the times to draw a single box plot.

(b) Explain why this is **not** an appropriate thing to do.

as he is comparing both schools so a single box won't help

(1)

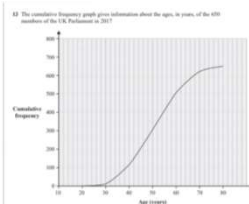
B1 e.g. in order to compare you would need to draw two separate box plots – one for Green Park school and one for Golden Plains school

B1 for identifying that more than one box plot would be required to allow for comparison

(1)

55

F P1 Q13b



(a) Using the cumulative frequency diagram, find an estimate of the median age for the members of the UK Parliament in 2017.

In 2017 the German Parliament had 55.4% of members of Parliament aged between 45 and 60 years old.

(b) Compare this figure to the percentage of members of the UK Parliament with ages between 45 and 60 years in 2017.

(b) Compare this figure to the percentage of members of the UK Parliament with ages between 45 and 60 years in 2017

$$500 - 215 = \frac{285}{600} = 0.475 \quad 47.5\%$$

M1 '510' and '210'

A1ft '300'

A1ft for '46.1538... %'

B1ft for percentage of MPs (aged between 45 and 60) was greater in Germany

M1 For their UB and their LB where

$$500 \leq UB \leq 520$$

$$200 \leq LB \leq 220$$

Note: check graph for indicative guidelines drawn within these ranges for UB and LB.

A1ft for their '300' (UB - LB)

$$280 \leq UB - LB \leq 320$$

A1ft for '46.1538... %' ~~avg~~

'46.2% Allow for a percentage which is correct to 1 decimal place.

B1ft for comparing percentages Follow through their percentage for UK provided M1 scored.

The German parliament had more members aged between 45 and 60 years old.

(4)

56

Foundation / Higher Crossover P2



57

F P1 Q13b

10 Matthew is investigating average household income for different states in the USA.

(a) Give a reason why it is appropriate to use secondary data for this.

because it's almost impossible to do by yourself

(1)

B1 for one of:

- it would be faster or it would take too long to collect the data himself
- data is easily accessible/easier to collect or too much data to collect/analyse / not practical to collect himself
- it would be cheaper or it would be too expensive to collect the data himself

B1 for identifying why it is appropriate to use secondary data.

Allow a disadvantage of primary data as long as it is clear that they are referring to primary data.

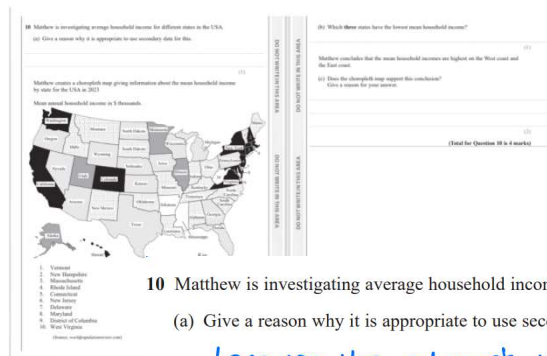
e.g. people might not want to tell Matthew their average income – sensitive question

B0 for more accurate or more data.

(1)

58

F P1 Q10a



10 Matthew is investigating average household income for different states in the USA.

(a) Give a reason why it is appropriate to use secondary data for this.

because it's almost impossible to do by yourself

(1)

B1 for one of:

- it would be faster or it would take too long to collect the data himself
- data is easily accessible/easier to collect or too much data to collect/analyse / not practical to collect himself
- it would be cheaper or it would be too expensive to collect the data himself

B1 for identifying why it is appropriate to use secondary data.

Allow a disadvantage of primary data as long as it is clear that they are referring to primary data.

e.g. people might not want to tell Matthew their average income – sensitive question

(1)

B0 for more accurate or more data.

59

F P1 Q11c

(c) Compare the three distributions of hand spans.

Give **three** comparisons and interpret **two** of your comparisons.

- International has the best range
- Amateur has the highest range
- International has the highest greatest hand span
- Amateur has the lowest greatest hand span

B1ft B1ft B1ft B1ft B1ft

Comparison	Interpretation
International greatest <u>median</u> or International <u>median</u> > national <u>median</u> > amateur <u>median</u>	e.g. International pianists have the largest/wider (hand spans). As they increase in standard the hand spans increase.
Amateur has the greatest <u>IQR</u> or International <u>IQR</u> < national <u>IQR</u> < amateur <u>IQR</u> or Amateur has the greatest <u>range</u> or	e.g. International pianists have <u>are</u> the most consistent

(2 2 2 2 2 2 2)

B1ft Correct medians

B1ft Correct comparison of spread (IQR or range)

B1ft Correct comparison of skew

B1ft One correct interpretation

B1ft One further correct comparison of spread or interpretation

Allow equivalent/converse statements but underlined words must be seen.

International range > national range < amateur range
or
All three have positive skew
or
International negative skew, national and amateur positive skew

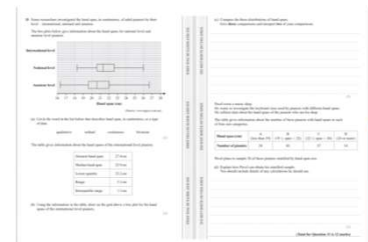
e.g. All three have more varied (hand spans) above median

Allow for comparison of just two box plots e.g. national and amateur

May be multiple comments in one statement.

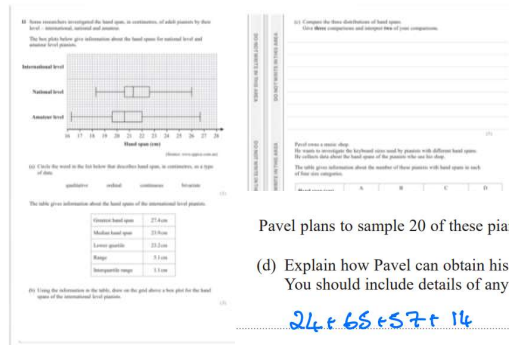
Follow through from their box plot.

Note: in this question ignore any



60

F P1 Q11d



Pavel plans to sample 20 of these pianists stratified by hand span size.

- (d) Explain how Pavel can obtain his stratified sample.
You should include details of any calculations he should use.

$$24 + 65 + 57 + 14 = 160 = 20 \times 8$$

$$\frac{24}{8} = 3 \quad \frac{65}{8} = 8$$

$$\frac{57}{8} = 7 \quad \frac{14}{8} = 2$$

B1 for e.g. $\frac{24}{24+65+57+14} \times 20$ or $\frac{\text{strata size}}{\text{total}} \times 20, \frac{1}{8}$ of each strata

B1 for one correct rounded value from 3, 8, 7 or 2
or indicating that they should round the sample size to the nearest whole number.

B1 for indicating or describing taking a random sample within each strata e.g. number all of the pianists and use a random number generator to select the appropriate number within the strata.

B1 for description of how to calculate the number to be sampled from each stratum
B1 for one correct integer value, ignore subsequent incorrect values, e.g. 3.8, 7.3 or indicating that they need to round the sample size to the nearest whole number.

B1 for indicating random sampling within each stratum or for description of how to sample within each stratum

(3)

61

Higher P2

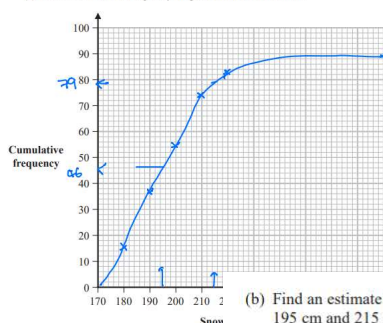
 **Pearson**
Edexcel



62

H P2 Q4b

(a) Draw a cumulative frequency diagram for this information.



(b) Find an estimate for the number of these days where the snow depth was between 195 cm and 215 cm.

$$\begin{array}{r} 79 \\ - 46 \\ \hline 33 \end{array}$$

33

(2)

M1 for 80 – 46

A1ft 34

segments.

Follow through their cumulative frequency graph.

Accept an answer in the range 31 – 37

Follow through an answer using their 80 and their 46 from a cumulative frequency graph.

(2)

63

H P2 Q5a

Welsh	England
8	7
7 6 4 4 4	8 8 7 4 2
8 6 5 5 2 1	8 6 5 4 3 2 0
9 4 0	7
	11 2 1

Key: 8 | 8 1 8 8

(Source: www.wales/fixtures-and-teams/teams/wales/)B2 for correctly completing the England Rugby Union player weights on the stem and leaf diagram
B1 for a suitable key

Welsh	England
8	7
7 6 4 4 4	8 2 4 7 8 8
8 6 5 5 2 1	9 0 2 3 4 6 6 6 8
9 4 0	10 7
	11 1 2

Key: 1 | 9 | 0 represents a weight of 91kg for a Welsh Rugby Union Back and a weight of 90kg for an England Rugby Union Back

B2 for a fully correct back-to-back stem and leaf diagram

OR if B2 not earned
B1 for unordered diagram or ordered diagram with at most 2 errorsAND
B1 for a suitable key for the stem and leaf diagram.
Accept a key given as two parts. If key given in two parts then this must be complete and there must be reference to Welsh and England or it must be clear how this is interpreted for the two sides. E.g. 8 | 7 represents a weight of 78 in Welsh and 8 | 2 represents a weight of 82 in England or | 8 | 2 represents 82, 4 | 8 | represents 84.

(3)

64

H P2 Q5a

Welsh	England
8	7
7 6 4 4 4	8 <i>24 788</i>
8 6 5 5 2 1	9 <i>0234 6668</i>
9 4 0	10 <i>7</i>
	11 <i>12</i>

Key: *8|2 = 82* *4|8 = 84*

(Source: www.wru.wales/fixtures-and-teams/teams/wales/)

B2 for correctly completing the England Rugby Union player weights on the stem and leaf diagram
B1 for a suitable key

Welsh	England
8	7
7 6 4 4 4	8 2 4 7 8 8
8 6 5 5 2 1	9 0 2 3 4 6 6 6 8
9 4 0	10 7
	11 1 2

Key: 1 | 9 | 0 represents a weight of 91kg for a Welsh Rugby Union Back and a weight of 90kg for an England Rugby Union Back

B2 for a fully correct back-to-back stem and leaf diagram

OR if B2 not earned
B1 for unordered diagram or ordered diagram with at most 2 errors

AND
B1 for a suitable key for the stem and leaf diagram.
Accept a key given as two parts.
If key given in two parts then this must be complete and there must be reference to Welsh and England or it must be clear how this is interpreted for the two sides. E.g. 8 | 7 represents a weight of 78 in Welsh and 8|2 represents a weight of 82 in England or |8|2 represents 82, 4|8| represents 84.

(3)

65

H P2 Q5 e/f

Amy wants to use the median and interquartile range statistics in a news article for a sports magazine. The article will compare the players on the two teams who are Backs.

(e) Comment on the appropriateness of using the median and the interquartile range in the article.

people might not understand what it means

(1)

(f) Give a limitation of using Zack's statistics to compare **all** the players on the two teams.

it doesn't show the heaviest or lightest

(1)

- (e) B1 for e.g.
- Not appropriate as readers may not understand what they (median and interquartile range) are
 - Appropriate if the meaning of the statistics were explained for the target audience

B1 for assessing the appropriateness of the use of median and interquartile range for the article
Allow not appropriate as IQR and median do not use all the data or appropriate as median and IQR are not affected by outliers.

(1)

- (f) B1 for e.g. the data is only for the Backs / we have no data for the Forwards / no data for all the positions / doesn't include all the players.

B1 for identifying a limitation of using these statistics to compare the two teams.
Allow not representative.
B0 data is only for two teams.
B0 reference the disadvantages of using the median and/or IQR.

(1)

66

H P2 Q6 b/c

surface runoff for a year.

- (b) Explain how Elizabeth's plan to collect data controls some extraneous variables. You should include in your answer an example of an extraneous variable that is likely to be controlled in this investigation.

she is measuring the same area and she is measuring for a year as well

(2)

Elizabeth would like to reduce the time that she is collecting data to one year overall.

- (c) Describe how she could do this using a matched pairs approach.

she can install the drainage solution in one area and compare it to another area

(1)

(b)	<p>B1 for e.g.</p> <ul style="list-style-type: none"> The same locations are tested without the drainage solution and then with the drainage solution <p>B1 for e.g.</p> <ul style="list-style-type: none"> Different terrain. Different surface materials e.g. tarmac. Differences in forest cover area. Different climate. <p>SCB2 for testing for a year without the drainage solution and for a year with the drainage solution controls for seasons / time of year</p>	B1 for identifying how the plan controls for extraneous variables.	(2)
(c)	<p>B1 for e.g.</p> <ul style="list-style-type: none"> Identify areas which have the same profiles e.g. same amount of rain, similar terrain. 	<p>B1 for describing a matched pairs approach where two areas are matched</p> <p>Allow for reference to two areas and matching of a feature.</p> <p>B0 for a description of testing the same area with drainage solution and then testing the same area after drainage solution.</p>	(1)

Table 1: Annual average CPI from 2017 to 2021 with 2015 as the base year.

Year	2017	2018	2019	2020	2021
Annual average CPI	103.6	106.0	107.8	108.9	111.6

(Source: www.ons.gov.uk)

67

H P2 Q7a

The Consumer Price Index (CPI) is a measure of the rate of change of prices in everyday life.

The table shows the annual average CPI from 2017 to 2021 with 2015 as the base year.

Year	2017	2018	2019	2020	2021
Annual average CPI	103.6	106.0	107.8	108.9	111.6

(Source: www.ons.gov.uk)

- (a) Give an interpretation of the number 108.9 in the table.

8.9% increase

<p>B2 CPI has increased 8.9% (from 2015 to 2020) (B1 for increase or 8.9%)</p>	<p>B2 for a complete interpretation of the CPI (B1 for increase or 8.9%)</p>	<p>(2)</p>
------------------------------------------------------------------------------------	----------------------------------------------------------------------------------	------------

68

H P2 Q7e

- (e) By calculating an appropriate geometric mean, compare the average annual inflation for Germany from 2017 to 2021 with the average annual inflation for the UK from 2017 to 2021.
You must show your working.

The table shows the annual average CPI from 2017 to 2021 for the UK and Germany.					
Year	2017	2018	2019	2020	2021
UK	102.56	102.29	101.74	100.99	102.52
Germany	102.56	102.29	101.74	100.99	102.52

all 5 marks are available for the answer (M1) to the table.

 The table gives the average CPI for the UK from 2017 to 2021. | | | | | | | |---------|--------|--------|--------|--------|--------| | Year | 2017 | 2018 | 2019 | 2020 | 2021 | | UK | 102.56 | 102.29 | 101.74 | 100.99 | 102.52 | | Germany | 102.56 | 102.29 | 101.74 | 100.99 | 102.52 | all 5 marks are available for the answer (M1) to the table. |

$$\sqrt[5]{2.56 \times 2.29 \times 1.74 \times 0.99 \times 2.52} = 1.91\%$$

The average annual inflation of the UK is greater than Germany

(3)

M1 for $\sqrt[5]{1.0256 \times 1.0229 \times 1.0174 \times 1.0099 \times 1.0252}$

A1 for 1.02018... [=2.02%]

OR

SCB1 for 1.91 or 1.91%

B1dep ft for average annual inflation rate was greater for the UK than for Germany (from 2017 to 2021)

Must see working to award M, A marks (arithmetic mean is also 2.02%)
A0 for an answer 1.02%

(3)

B1depft on one previous mark scored.



Pearson