GCSE Statistics

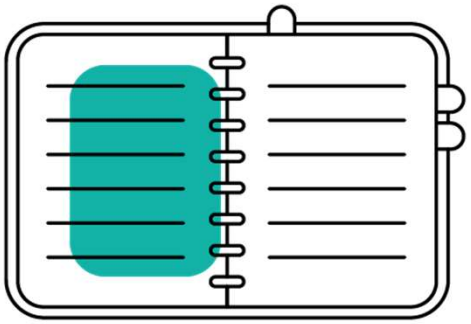
Marking Guidance Summer 2024

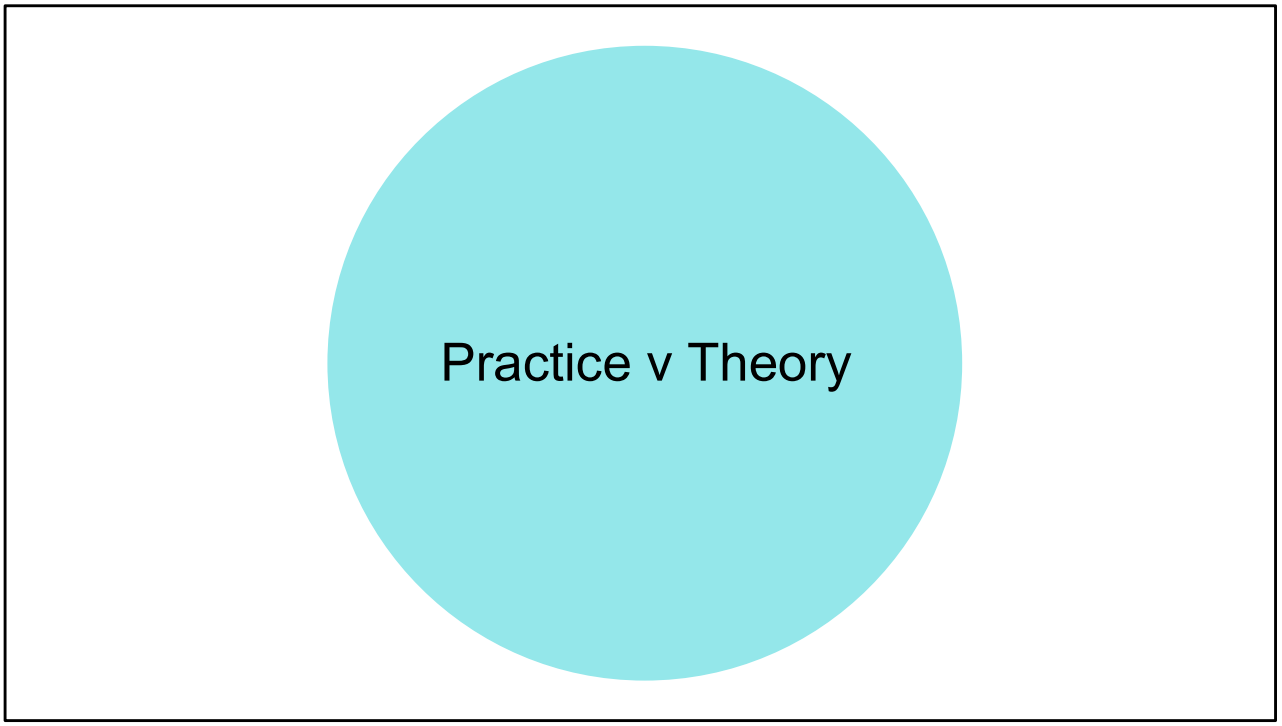
Mel Muldowney
Christian Seagar

JustMaths

Agenda

- General marking guidance
- Quick paper overview
- Specific questions
  - Foundation
  - Crossover
  - Higher
- Note: extended answers … more training coming soon!




Practice v Theory

General … it is different to GCSE Maths?

Guidance on the use of abbreviations within this mark scheme	
M	method mark awarded for a correct method or partial method
A	accuracy mark (awarded after a correct method; if no method is seen then full marks for the question are implied but see individual mark schemes for more details)
B	unconditional accuracy mark (no method needed)
oe	or equivalent
cao	correct answer only
ft	follow through (when appropriate as per mark scheme)
sc	special case
dep	dependent (on a previous mark)
indep	independent
awrt	answer which rounds to
isw	ignore subsequent working

General marking principles

General marking guidance

These notes offer general guidance, but the specific notes for examiners appertaining to individual questions take precedence.

1 All candidates must receive the same treatment. Examiners must mark the last candidate in exactly the same way as they mark the first.

Where some judgement is required, mark schemes will provide the principles by which marks will be awarded; exemplification/indicative content will not be exhaustive. When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the response should be sent to review.

2 All the marks on the mark scheme are designed to be awarded; mark schemes should be applied positively. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme. If there is a wrong answer (or no answer) indicated on the answer line always check the working in the body of the script (and on any diagrams), and award any marks appropriate from the mark scheme.

Questions where working is not required: In general, the correct answer should be given full marks.
Questions that specifically require working: In general, candidates who do not show working on this type of question will get no marks – full details will be given in the mark scheme for each individual question.

3 **Crossed out work**
This should be marked **unless** the candidate has replaced it with an alternative response.

4 **Choice of method**
If there is a choice of methods shown, mark the method that leads to the answer given on the answer line.

If no answer appears on the answer line then mark both methods **as far as they are identical** and award these marks.

5 **Incorrect method**
If it is clear from the working that the "correct" answer has been obtained from incorrect working, award 0 marks.

General marking principles

6 Follow through marks

Follow through marks which involve a single stage calculation can be awarded without working as you can check the answer, but if ambiguous do not award.
Follow through marks which involve more than one stage of calculation can only be awarded on sight of the relevant working, even if it appears obvious that there is only one way you could get the answer given.

7 Ignoring subsequent work

It is appropriate to ignore subsequent work when the additional work does not change the answer in a way that is inappropriate for the question or its context. (eg an incorrectly cancelled fraction when the unsimplified fraction would gain full marks).
It is not appropriate to ignore subsequent work when the additional work essentially makes the answer incorrect (eg incorrect algebraic simplification).

8 Probability

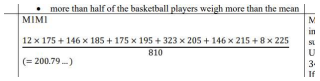
Probability answers must be given as a fraction, percentage or decimal. If a candidate gives a decimal equivalent to a probability, this should be written to at least 2 decimal places (unless tenths).
Incorrect notation should lose the accuracy marks, but be awarded any implied method marks.
If a probability fraction is given then cancelled incorrectly, ignore the incorrectly cancelled answer.

9 Range of answers

Unless otherwise stated, when an answer is given as a range (eg 3.5 – 4.2) then this is inclusive of the end points (eg 3.5, 4.2) and all numbers within the range.

General … it is different to GCSE Maths?

NOT EXPLICIT IN THE MARKSCHEMES



• more than half of the basketball players weigh more than the mean	
M1M1	
$\frac{12 \times 175 + 146 \times 185 + 175 \times 195 + 323 \times 205 + 146 \times 215 + 8 \times 225}{810}$	
(= 200.79 ...)	

11 **Number in brackets after a calculation**
Where there is a number in brackets after a calculation E.g. 2 × 6 (=12) then the mark can be awarded **either** for the correct method, implied by the calculation **or** for the correct answer to the calculation.

12 **Use of inverted commas**
Some numbers in the mark scheme will appear inside inverted commas E.g. "12" × 50 ; the number in inverted commas cannot be any number – it must come from a correct method or process but the candidate may make an arithmetic error in their working.

13 **Word in square brackets**
Where a word is used in square brackets E.g. [area] × 1.5 : the value used for [area] does **not** have to come from a correct method or process but is the value that the candidate believes is the area. If there are any constraints on the value that can be used, details will be given in the mark scheme.

14 **Misread**
If a candidate misreads a number from the question. Eg. uses 252 instead of 255; method or process marks may be awarded provided the question has not been simplified. Examiners should send any instance of a suspected misread to review.

Pearson Edexcel

Foundation P1

Skills Map

Q	Part	Skill tested	Mean score	Max score	Mean %	5	4	3	2	1	U	ALL
Q01	a	Pictogram – Completing	0.93	1	93	0.98	0.98	0.97	0.92	0.81	0.51	0.93
Q01	b	Pictogram – Using	1.73	2	87	1.92	1.87	1.82	1.68	1.35	0.64	1.73
Q01	c	Pictogram – Comparing	1.20	2	60	1.47	1.36	1.26	1.08	0.75	0.27	1.20
Q01	d	Importance of reliability and validity	1.16	2	58	1.67	1.47	1.18	0.82	0.43	0.11	1.16
Q02	ai	Bar charts – completing	0.79	1	79	0.97	0.94	0.85	0.69	0.38	0.07	0.79
Q02	aii	Bar charts – completing	0.50	1	50	0.78	0.65	0.48	0.33	0.15	0.02	0.50
Q02	b	Bar charts – completing	1.51	2	76	1.88	1.77	1.62	1.36	0.80	0.16	1.51
Q02	c	Bar charts – Comparing	0.99	2	50	1.58	1.25	0.94	0.67	0.34	0.05	0.99
Q02	d	Justify the appropriateness	0.39	2	20	0.85	0.49	0.29	0.16	0.05	0.01	0.39
Q03	ai	Tabulation – Reading	0.90	1	90	0.97	0.94	0.92	0.89	0.80	0.42	0.90
Q03	aii	Tabulation – Reading	0.83	1	83	0.93	0.90	0.86	0.80	0.65	0.26	0.83
Q03	b	Tabulation – Interpreting	0.65	1	65	0.84	0.75	0.65	0.56	0.39	0.10	0.65
Q03	c	Tabulation – Comparing	1.50	2	75	1.77	1.70	1.58	1.39	1.01	0.28	1.50
Q03	d	Time series	0.84	2	42	1.35	1.08	0.84	0.52	0.16	0.01	0.84
Q03	e	Identify trends in data	0.74	1	74	0.90	0.86	0.80	0.67	0.39	0.09	0.74
Q04	a	Definitions of different types of data	0.34	1	34	0.53	0.39	0.31	0.24	0.20	0.10	0.34
Q04	b	Calculate median	1.49	2	75	1.84	1.74	1.59	1.35	0.80	0.13	1.49
Q04	c	Justify the rationale for using median	0.19	1	19	0.32	0.21	0.16	0.14	0.09	0.03	0.19
Q04	d	Calculate range	1.61	2	81	1.97	1.93	1.80	1.40	0.73	0.09	1.61
Q04	e	Compare data sets using range and median	1.68	4	42	3.17	2.32	1.52	0.70	0.14	0.01	1.68
Q05	a	Sample space diagrams – completing	1.49	2	75	1.92	1.80	1.57	1.24	0.70	0.22	1.49
Q05	bi	Sample space diagrams – using	0.53	1	53	0.89	0.73	0.51	0.28	0.09	0.01	0.53
Q05	bii	Sample space diagrams – using	0.26	1	26	0.61	0.34	0.17	0.08	0.03	0.00	0.26
Q05	c	Experimental v theoretical values	1.09	2	55	1.70	1.40	1.11	0.71	0.24	0.04	1.09
Q06	a	Hypothesis testing	0.37	1	37	0.69	0.50	0.33	0.16	0.07	0.01	0.37
Q06	b	Systematic sampling	0.12	1	12	0.31	0.17	0.07	0.03	0.00	0.00	0.12
Q06	ci	Systematic sampling	0.42	1	42	0.69	0.56	0.39	0.23	0.09	0.01	0.42
Q06	cii	Systematic sampling	0.40	2	20	1.02	0.58	0.22	0.05	0.02	0.00	0.40
Q06	d	Questionnaires	0.73	2	37	1.05	0.89	0.76	0.57	0.28	0.06	0.73
Q06	ei	Planning data collection	0.12	1	12	0.29	0.15	0.07	0.03	0.01	0.00	0.12
Q06	eii	Planning data collection	0.87	2	44	1.39	1.13	0.84	0.56	0.25	0.03	0.87
Q06	f	Reliability and validity	0.60	1	60	0.80	0.73	0.64	0.48	0.25	0.04	0.60
Q07	a	Pie chart – comparing	1.28	2	64	1.66	1.55	1.39	1.05	0.52	0.10	1.28
Q07	b	Pie chart	0.54	2	27	1.29	0.76	0.33	0.10	0.02	0.00	0.54
Q08	a	Tabulation – Reading	0.84	2	42	1.56	1.19	0.74	0.32	0.07	0.00	0.84
Q08	bi	Arithmetic mean	0.69	3	23	1.73	0.91	0.43	0.16	0.04	0.00	0.69
Q08	bii	Arithmetic mean	0.41	1	41	0.77	0.56	0.36	0.17	0.05	0.01	0.41
Q08	biii	Arithmetic mean	0.33	1	33	0.68	0.46	0.27	0.12	0.04	0.01	0.33
Q08	c	Select and justify appropriate diagrams	0.13	2	7	0.40	0.15	0.05	0.02	0.01	0.00	0.13
Q09		Population pyramid	1.27	5	25	2.65	1.68	1.05	0.48	0.11	0.00	1.27
Q10	a	Histograms– using	1.56	2	78	1.94	1.88	1.74	1.34	0.66	0.14	1.56
Q10	b	Histograms – using	0.16	2	8	0.46	0.17	0.08	0.06	0.03	0.01	0.16
Q10	c	Skewness	0.11	2	6	0.29	0.14	0.08	0.03	0.01	0.00	0.11
Q11	a	Tree diagrams	1.14	2	57	1.82	1.54	1.15	0.64	0.16	0.02	1.14
Q11	b	Formal notation for independent events	1.10	4	28	3.10	1.47	0.50	0.16	0.02	0.01	1.10
			36.53	**80**	**46**	**56.40**	**45.04**	**35.29**	**25.44**	**14.19**	**4.08**	**36.53**

Examiners Reports – F P1

Students were generally able to attempt the whole paper within the time allowed. Students performed well on the first five questions, and there were varying degrees of success on most of the remaining questions. Question 6 proved to be more challenging for candidates to with many not using statistical processes correctly. Most candidates were able to access some early parts of these later questions and pick up valuable marks.

Based on their performance on this paper, candidates should:

- read each question fully and carefully before attempting to answer it;
- show working out to support the final answer;
- ensure that they are able to give comparisons as part of a response;
- practice interpreting answers to statistical calculations and diagrams in context;
- practice identifying the appropriateness of calculations, diagrams and approaches;
- give a decision when the question asks for this and support it using the information from the question;
- practice more extended response questions and understand how to break these down with the steps outlined in the question;
- review key terms and processes as part of the statistical cycle, particularly those exemplified in question 6: hypotheses, pilot studies, improving and writing questionnaires and benefits of sampling.

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

70

(2)

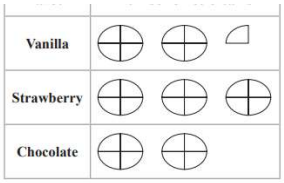
number		
1(a)	B1 For two and a half images drawn.	Accept half circle with no defined quarters. Ignore size of circles.
(b)	M1 $(2 \times 8 + \frac{1}{4} \times 8) + (3 \times 8) + 20$ or $(7 \times 8) + (\frac{3}{4} \times 8)$ A1 62	M1 for $18 + 24 + 20$ Allow one incorrect ice cream flavour total from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

(b) M1A0

(b) The correct sum is seen that has been evaluated incorrectly so the method mark is scored

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.



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.

(c) B1

The stated values are correct but there is not a comparison.

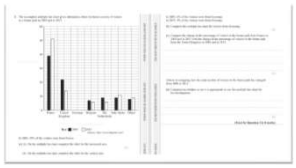
(d) B2

A correct conclusion and their reason is highlighting a large variation in data (point 5 on the mark scheme).

F P1 Q2c/d

(c)	B1B1 for each correct comparison • 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%	Allow for B1B1 France has increased whereas United Kingdom has decreased If they have stated 'increase' or 'decrease' correctly then ignore figures If they have stated 'changed' then figures must be correct Condone reference to number instead of percentage. B0 for comparison of France v UK in 2003 or for comparison of France v UK in 2013

(d)	B2 for a correct reason **and** conclusion e.g. • 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. (B1 for e.g. • The graph only shows percentages • The graph does not show the total number of visitors • The graph only shows 2003 and 2013)	B2 for not appropriate with a correct reason (B1 for a correct reason and no conclusion)



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
2003 = 39% % of visitors
2013 = 51% increased

U.K.
2003 22% % of visitors
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.

(c) B2

(c) There are two correct comparisons and so the marks are awarded. The percentages stated are correct. These are not required, nor is a numerical difference for these values needed. In most instances a conclusion can be drawn from comparing the difference in heights.

(d) B0

(d) An incorrect response. Any response that starts 'It is appropriate' is incorrect.

F P1 Q2c/d

(c)	B1B1 for each correct comparison • 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%	Allow for B1B1 France has increased whereas United Kingdom has decreased If they have stated 'increase' or 'decrease' correctly then ignore figures If they have stated 'changed' then figures must be correct Condone reference to number instead of percentage. B0 for comparison of France v UK in 2003 or for comparison of France v UK in 2013

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)

(d)	B2 for a correct reason **and** conclusion e.g. • 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. (B1 for e.g. • The graph only shows percentages • The graph does not show the total number of visitors • The graph only shows 2003 and 2013)	B2 for not appropriate with a correct reason (B1 for a correct reason and no conclusion)

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)

(c) B2

(c) A correct response that states that the change in visitors increased in France and decreased in the UK - both marks are awarded for this. There is no need to quantify this change. The candidate has also stated, unnecessarily, that the change is bigger in France - which is correct and improves their response further.

(d) B0

(d) An incorrect response.

F P1 Q2c/d

(c)	B1B1 for each correct comparison • 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%	Allow for B1B1 France has increased whereas United Kingdom has decreased If they have stated 'increase' or 'decrease' correctly then ignore figures If they have stated 'changed' then figures must be correct Condone reference to number instead of percentage. B0 for comparison of France v UK in 2003 or for comparison of France v UK in 2013

(d)	B2 for a correct reason **and** conclusion e.g. • 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. (B1 for e.g. • The graph only shows percentages • The graph does not show the total number of visitors • The graph only shows 2003 and 2013)	B2 for not appropriate with a correct reason (B1 for a correct reason and no conclusion)

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 14 percent.

(2)

ohn is investigating how the total number of visitors to the theme park has changed om 2003 to 2013

) 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 chat

(2)

(c) B0

(c) There is no comparison so no marks. There are no marks awarded for stating the percentages from the graph.

(d) B0

(d) An incorrect response.

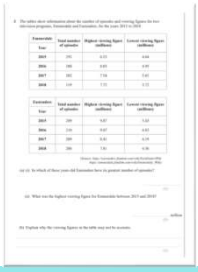
F P1 Q3b

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

people could be watching from websites instead of TV

(1)

(b)	B1 for e.g. • 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	B1 for correct reason why the figures may be inaccurate



B1 Could not measure the number of people accurately

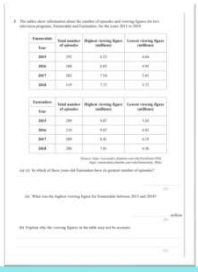
F P1 Q3b

(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)

(b)	B1 for e.g. • 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	B1 for correct reason why the figures may be inaccurate



B1 Reference to rounding

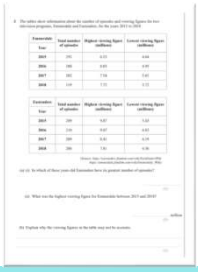
F P1 Q3b

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

could be secondary data

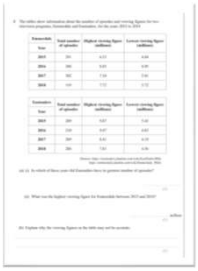
(1)

(b)	B1 for e.g. • 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	B1 for correct reason why the figures may be inaccurate



B1 Mentions that data is secondary data. (bullet point 7)

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)	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

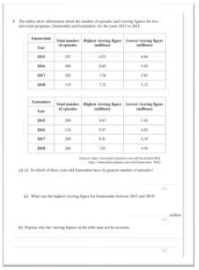
B1 - 1 mark Although a minimal response, this clearly articulates the trend and the mark is awarde

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

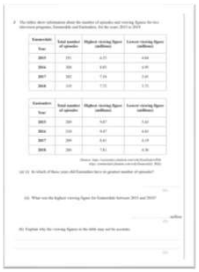
(2)



(c)	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

B0 - no marks This question is testing trend and not correlation and so, as stated in additional guidance, this is not accepted.

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 decreased from 15 to 13 from 9.87 million to 7.81 million

(2)

(c)	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

B1 - 1 mark The mark is awarded for "decreases". Any adjective, "slowly", used can be ignored. N.B Any quoted figures can be ignored, as this question is testing the candidate's understanding of trend.

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 visitor views down

(2)

(c)	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

B1 - 1 mark Although the response begins with "negative correlation", which is not accepted, the correct description for the trend is then seen. N.B We always try to mark positively and will condone the inclusion of correlation in this response.

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) • Range of the first 9 games is bigger (than the range of the final 9 games)	e.g. • On average they score more points at the end of the season • 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

ft their median in part (b) and their range in (d)

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

e) 4 marks.

B1 - comparison of 'their' median with 90 and they have said that it has increased.

B1 - 'In the final 9 matches the team has become better' is their interpretation of their medians and scores the mark.

B1 - correct comparison of the range - range has decreased from 96 to 25

B1 - correct interpretation - more consistent, best & worst scores not as far apart.

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: e.g. • Median of first 9 games is smaller (than the median of the final 9 games) • Range of the first 9 games is bigger (than the range of the final 9 games) Interpretation: e.g. • On average they score more points at the end of the season • 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 ft their median in part (b) and their range in (d) Allow equivalent/converse statements but underlined words must be seen.

e) 3 marks

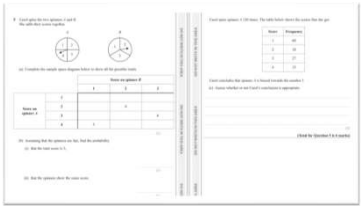
B1 - median of final matches higher

B1 - better average score

B1 - higher range for 1st 9 games

B0 - bigger outliers is not a sufficient interpretation.

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.

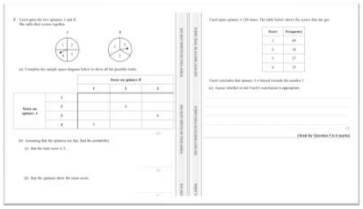
Yes it is appropriate as the number 1 got much more than the rest and therefore is biased.

(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

B1 B1 - conclusion is appropriate with correct reason (bullet point 1)

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

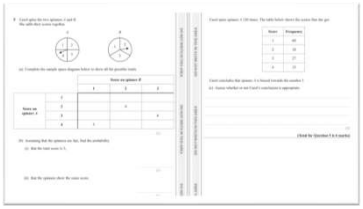
(2)

(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

B1 - half the spins landed on 1 (bullet point 1)

B1 - it is a fair assessment implies appropriateness/Carol is correct

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 as the pattern is 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

B1 - would expect all scores to be similar (bullet point 4)

B0 - no mention of appropriateness

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
	B1 Systematic	

a) B0 - 0 marks,

b) (a) This is not a hypothesis. It is a reason as to why perhaps the gym should open longer.

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.

		Ignore additional non-contradictory comments
(e)(i)	B1 Pilot	Accept pilot study or pilot test or pilot survey
(ii)	B1B1 for each of two from e.g. • 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

(e i) B0 - 0 marks,

(e i) Incorrect.

(e ii) B1 -- 1 mark,

(e ii) To predict the outcome is incorrect. The point of this question is about refining the questionnaire. The second reason offered "change the questionnaire" is a good response and scores.

F P1 Q7a/b

7(a)	B1B1 for each of two from e.g. • a greater proportion of houses were for sale in Harrogate oe • a lower proportion of flats were for sale in Harrogate oe • a similar proportion / a (slightly) lower proportion of bungalows were for sale in Harrogate and South Shields oe • a similar proportion / a (slightly) lower proportion of other properties were for sale in Harrogate oe **Harrogate** / **South Shields** **Houses**: Greater / Lower **Flats**: Lower / Greater **Bungalows**: Greater / Similar / Lower / Similar **Other**: Lower / Similar / Greater / Similar	Condone the omission of the word proportion or the use of number for proportion Allow the use of angles / sectors Allow converse. Ignore additional non-contradictory comments Allow B1 for e.g. Houses are greatest in both
(b)	B2 not valid and a reason given • Since the pie charts only show proportions and not numbers • Totals may be different / do not know exact amounts OR B2 may be valid and a reason given • If the same number of properties are for sale in both Harrogate and South Shields. (B1 for a correct reason and no conclusion / incorrect conclusion)	

(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)

(a) B1

(b) (a) See additional guidance - B1 - Houses are greatest in both

(b) B2

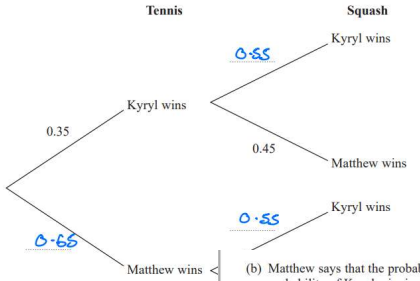
(b) A correct reason of populations and a conclusion stated - two marks.

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.

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

(4)

(Total for Question 11 is 6 marks)

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) depB1ft Matthew is correct (as 0.2925 > 0.1925)	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 dep on M1 and two probabilities calculated. Their probabilities must be 0<p<1

(a) B2 (b) M0A0A0B0

(b) (a) The tree diagram has been completed correctly.

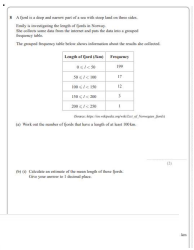
(c) (b) No marks can be scored unless probabilities are multiplied.

Pearson Edexcel

Foundation / Higher Crossover P1

F P1 Q8 bii / iii and c

(ii)	B1 for any one from • 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 • 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 • (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 • 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 times 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)

(bii) B0 - 0 marks,

(bii) Reference to secondary data does not explain why it is an estimate - no marks.

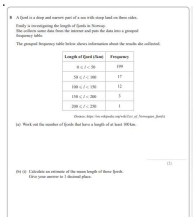
(biii) B0 - 0 marks,

(biii) Repeating the calculations or checking is a good technique but is not what we are testing and so no marks are awarded.

(c) B0 - 0 marks

(c) While graphs may shows patterns and trends - this does not explain why or why not frequency polygons are appropriate in this question. No marks.

F P1 Q8 b ii / iii and c



(ii)	B1 for any one from • 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 • 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 • (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 • 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 cant represent the data properly

(2)

(bii) B1 - 1 mark,

(bii) The use of "class widths" or "not the exact values" are both good enough alone to score the mark. Together this is a strong correct response.

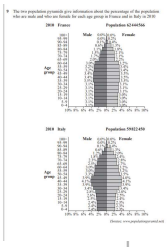
(biii) B1 - 1 mark,

(biii) Using exact values is equivalent to using raw/ungrouped data and scores the mark.

(c) B0 - 0 marks

(c) There is no reason offered as to whether it is appropriate to use a frequency polygon. No marks.

F P1 Q9



y is investigating how the populations of Italy and France differ in 2010
:s the two population pyramids to reach the following two conclusions.

e percentage of people aged 50–54 was lower in France than the percentage of
ople aged 50–54 in Italy.

e number of males aged 40–44 in France was greater than the number of males
ed 40–44 in Italy.

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

France
50-54 3.2%
40-44 3.4%.
3.4 of 62444566
= 2 12845

Italy
50-54 = 3.8%.
40-44 = 4.1%
4.1% of 59822450
= 2452720

The percentages of people aged 50-54 in France was lower than the percentag of people aged 50-54 in Italy as France had 3.2%. whilst Italy had 3.8%. Tommys first conclusion was right, however for the number of males aged 40-44 in France was not greater than the numbe of males in Italy as Italy had 329605 more males aged 40-44 than France.

(Total for Question 9 is 5 marks)

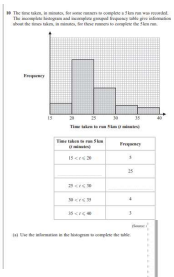
M1 3.2 + 3.4 = (6.6) or 3.3 + 3.5 = (6.8) A1 for 6.6 and 6.8	M1 for identifying (France) 3.2, 3.4 AND (Italy) 3.3, 3.5 A1 for 3.2 < 3.3 AND 3.4 < 3.5	M1 for attempt at finding either total percentage or for identifying all four required percentages A1 for both correct totals or for both correct comparisons
M1 for 0.034 × 62 444 566 = (2 123 115.244) or 0.041 × 59822450 = (2 452 720.45) A1 for 2 123 115 and 2 452 720		M1 for correct method for finding the number of males aged 40-44 or the number of males aged 40-44 in Italy A1 for both correct. Condone if not integers
dep B1ft Claim one is correct / the percentage of 50-54 is greater in Italy compared to France. **AND** claim two is incorrect / the number of males aged 40-44 is greater in Italy compared to France.		dep B1ft for accepting claim one and rejecting claim two. Dependent on M1M1

M0A0 They did not calculate either total percentage for 50-54, and these are not on the diagram either. They did read off the percentages for males aged 50-54, but they would need to read off all four percentages (males and females for each country) in order to get this M1 if they hadn't added to get totals.

M1A1 They have calculated the numbers of males aged 40-44. Correct rounded answers imply multiplication, so if only one value had been correct they would get M1A0

B0 This mark is dependent on M1M1. In t his case they got M0M1 so they can't achieve this B mark

F P1 Q10c



Time taken to run 5 km (t minutes)	Frequency
$15 < t \leqslant 20$	8
	25
$25 < t \leqslant 30$	
$30 < t \leqslant 35$	4
$35 < t \leqslant 40$	3

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

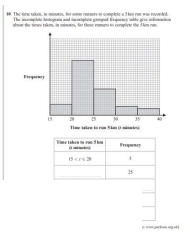
it has a positive correlation.

(2)

	A1 20	
(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

B0 - 0 marks Although the key word "positive" is seen, it is used in the context of correlation - which is wrong and scores no marks

F P1 Q10c



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

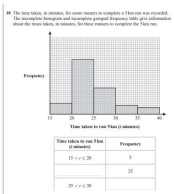
It is a positive skew as more people took 20 - 25 minutes

(2)

	A1 20	
(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

B1 - 1 mark The correct description of skew is seen and scores the first mark. Although there is a correct fact about the histogram this does not interpret the skew and does not score.

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)

	A1 20	
(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

B2 - 2 marks The correct skew is identified and then reference to the median being to the left of the diagram is the correct description of "positive" skew.

Higher P1

S

Q	Part	Skill tested	Mean score	Max score	Mean %	9	8	7	6	5	4	3	U	ALL
Q01	ai	Rates of change over time - using	1.85	2	93	1.98	1.96	1.93	1.89	1.81	1.63	1.28	0.97	1.85
Q01	aii	Rates of change over time - using	0.96	1	96	0.99	0.98	0.98	0.97	0.96	0.90	0.85	0.72	0.96
Q01	b	Rates of change over time - using	0.81	2	41	1.08	0.87	0.83	0.77	0.77	0.68	0.65	0.42	0.81
Q02	a	Tabulation - Reading	1.55	2	78	1.95	1.86	1.74	1.59	1.37	1.07	0.75	0.38	1.55
Q02	bi	Arithmetic mean	2.16	3	72	2.90	2.79	2.61	2.27	1.77	1.07	0.43	0.19	2.16
Q02	bii	Arithmetic mean	0.84	1	84	0.99	0.98	0.94	0.90	0.76	0.58	0.39	0.13	0.84
Q02	biii	Arithmetic mean	0.67	1	67	0.88	0.81	0.76	0.70	0.59	0.44	0.25	0.02	0.67
Q02	c	Select and justify appropriate diagrams	0.68	2	34	1.27	1.03	0.82	0.67	0.43	0.22	0.09	0.02	0.68
Q03		Population pyramid	3.30	5	66	4.69	4.31	3.84	3.33	2.61	1.91	1.24	0.48	3.30
Q04	a	Histograms- using	1.95	2	98	1.99	1.97	1.96	1.95	1.95	1.92	1.89	1.58	1.95
Q04	b	Histograms - using	1.00	2	50	1.80	1.61	1.31	0.96	0.55	0.25	0.16	0.06	1.00
Q04	c	Skewness	1.05	2	53	1.79	1.57	1.32	1.05	0.65	0.36	0.20	0.05	1.05
Q05	ai	Reasons for employing judgement sampling	0.26	1	26	0.65	0.45	0.32	0.21	0.11	0.06	0.03	0.00	0.26
Q05	aii	Reliability and validity	1.33	2	67	1.76	1.68	1.51	1.35	1.16	0.84	0.48	0.20	1.33
Q05	b	Questionnaires	1.03	2	52	1.65	1.42	1.22	0.97	0.78	0.55	0.31	0.16	1.03
Q05	c	Problems with collected data	0.98	2	49	1.31	1.16	1.06	0.99	0.86	0.74	0.51	0.23	0.98
Q05	d	Systematic sampling	0.90	3	30	1.96	1.46	1.11	0.80	0.49	0.26	0.10	0.02	0.90
Q05	e	Interviews v questionnaires	0.45	1	45	0.74	0.62	0.49	0.42	0.36	0.29	0.19	0.10	0.45
Q06	a	Identify trends in data	1.59	2	80	1.96	1.92	1.84	1.67	1.38	1.03	0.65	0.37	1.59
Q06	b	Interpret seasonal trends in context	2.00	3	67	2.66	2.46	2.28	2.04	1.78	1.26	0.51	0.08	2.00
Q06	ci	Interpret seasonal trends in context	0.44	2	22	1.63	1.12	0.47	0.16	0.03	0.01	0.00	0.00	0.44
Q06	cii	Interpret seasonal trends in context	0.18	1	18	0.73	0.46	0.17	0.05	0.01	0.00	0.00	0.00	0.18
Q06	di	Use trends to make predictions	0.61	2	31	1.90	1.55	0.81	0.25	0.05	0.01	0.00	0.00	0.61
Q06	dii	Use trends to make predictions	1.27	2	64	1.94	1.83	1.58	1.26	0.93	0.55	0.20	0.02	1.27
Q07	a	Sources of data	0.51	1	51	0.76	0.64	0.58	0.51	0.42	0.33	0.22	0.09	0.51
Q07	b	Hypothesis testing	0.82	1	82	0.99	0.97	0.93	0.86	0.75	0.53	0.28	0.12	0.82
Q07	c	Select and justify appropriate diagrams	1.22	3	41	2.48	2.09	1.67	1.09	0.56	0.25	0.07	0.05	1.22
Q07	d	Comparative pie chart	0.72	2	36	1.90	1.60	1.01	0.44	0.12	0.02	0.01	0.00	0.72
Q07	e	Comparative pie chart	0.57	1	57	0.82	0.77	0.69	0.57	0.43	0.31	0.14	0.05	0.57
Q08	a	Sample means a	0.61	2	31	0.99	0.77	0.63	0.58	0.50	0.41	0.27	0.06	0.61
Q08	b	Use action and warning lines i	1.70	2	85	1.94	1.87	1.82	1.73	1.64	1.43	1.01	0.41	1.70
Q08	c	Use action and warning lines i	1.55	5	31	4.19	3.47	2.21	0.95	0.25	0.07	0.03	0.00	1.55
Q08	d	Use action and warning lines i	0.32	1	32	0.59	0.45	0.35	0.28	0.23	0.19	0.19	0.14	0.32
Q09	a	Tree diagrams	2.23	4	56	3.58	3.07	2.62	2.17	1.65	1.10	0.69	0.49	2.23
Q09	b	Characteristics of a binomial distribution	0.87	4	22	3.26	2.07	1.01	0.35	0.09	0.03	0.01	0.00	0.87
Q10		Use means and standard deviation to standardise and interpret data	2.83	6	47	5.49	5.06	4.31	2.56	0.95	0.33	0.12	0.07	2.83
			41.81	**80**	**52**	**68.19**	**59.70**	**49.73**	**39.31**	**29.75**	**21.63**	**14.20**	**7.68**	**41.81**

Examiners Reports – H P1

Most candidates responded to the challenges within this paper well and demonstrated understanding of a range of areas of the specification. They were generally confident at completing calculations and diagrams and demonstrated good statistical understanding when asked to interpret these. As seen in previous series, candidates found questions requiring interpretation in context and evaluation of approaches or techniques more slightly more challenging.

Candidates should be encouraged to show full working and set this out clearly so that partial credit can be awarded if a fully correct solution is not obtained. They should also read the question carefully to identify the demand, for example whether an interpretation in context or conclusion is required.

Based on their performance on this paper, students should:

• show working for statistical calculations;

• practise writing clear explanations, bearing in mind exactly what is asked in the question and what evidence you should give to support your answer;

• practise interpreting statistical calculations in the context of the question;

• ensure that they can recognise and name different sampling techniques;

• ensure that they can describe how to take different types of sampling;

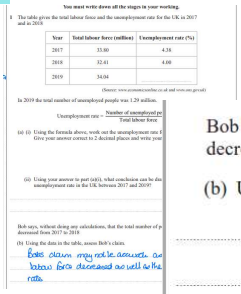
• practise giving statistical reasons or against the diagrams suggested;

• practise calculating mean seasonal variation from a time series graph and using this to make predictions;

• develop their understanding of control charts and the actions to be taken for samples falling in different parts of these charts;

• practise using the binomial distribution to calculate probabilities; • practise working with and interpreting standardised scores in a variety of contexts.

H P1 Q1b



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.

Bobs claim may not be accurate as the total labour force decreased as well as the unemployment rate

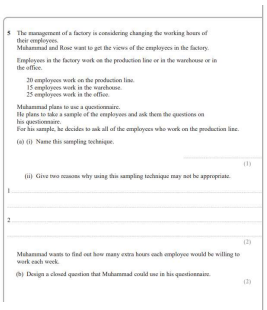
(2)

(Total for Question 1 is 5 marks)

(b)	B2 for a correct decision and complete reason e.g. • Bob is correct as the total workforce **and** unemployment rate is lower in 2018 compared to 2017 (B1 for e.g. • 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)	B2 for a complete assessment of the claim together with reason (B1 for a complete reason and no or incorrect conclusion or for a correct decision with partial reason). Not 1.29 million (2019 figure)

(b) B1 for identifying the correct reason but not having a correct conclusion

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.

1 .....

2 .....

(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.

(3)

(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 numbe generator choose 10% so 6 numbers and then you have your sample.

(3)

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.

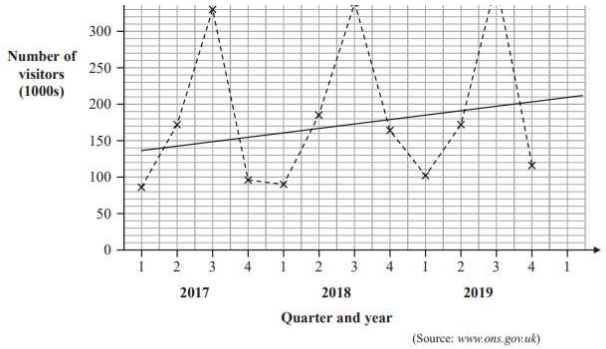
(1)

(d)	B1 get a list of all the 60 employees / get a sample frame / number all the employees B1 sample every 10th 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'.

d) B1 for number all 60 employees B0 for no mention of a systematic approach of selecting every 10th person

B0 no mention of a random start point e) B0 for no correct answer

H P1 Q6a



(Source: www.ons.gov.uk)

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 numbe 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.

a) B1 for rising trend B1 for correct contextual interpretation ‘

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 - 70
185 - 100 = 85

$\frac{50+70+85}{3}$

68·3333

68·3 thousand

(2)

(c)(i)	M1 $\frac{\pm(-50+-70+-85)}{3}$ 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**.

ci) M1 for correct calculation with incorrect signs A0 answer not given as -68.3'

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. • 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.

B1 "Respected" is equivalent to trustworthy. B1 Correct response.

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 shos there was more people in 2003

(3)

(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).	

B1 Reference has been made to pie charts showing proportions by writing "not just in proportion“

. B1 A correct reason "comparative pie charts..... more people in 2003" stated.

B0 as no decision has been reached.

H P1 Q10

	Answer	Additional guidance
	M1 $\frac{13.35-14.5}{2.3} = (-0.5)$ A1 -0.5	M1 for $\pm\left(\frac{13.35-14.5}{2.3}\right)$ A1 -0.5
	M1 $(0.3 \times 5.4) + 57.2\ (= 58.82)$ A1 58.8(2)	M1 for method to find time for Kai A1 58.8(2)
	dep B1ft for eg • Dominic is incorrect as his standardised score is lower for the 100 m • Dominic is incorrect as $-0.5 < -0.2$ oe.	DepB1ft for a correct conclusion with comparison of standardised scores for Dominic. Follow through their standardised score. Dep on first M scored Do not allow reference to standardised scores being closer to 1 or -1 or 0.
	depB1ft Kai is correct as his time is over one second longer (than Dominic's).	DepB1ft for a correct conclusion with comparison of times for 100 m. Dep on second M scored. Follow through their time for Kai.

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 100m

(Total for Question 10 is 6 marks)

M1A1 for correct calculation shown and evaluated for standardised score

M1A1 for correct calculation shown and evaluated for time

B1 B1 for two correct conclusions, with reasons, made regarding both statements.

Pearson Edexcel

Foundation P2

Skills M

Q	Part	Skill tested	Mean score	Max score	Mean %	5	4	3	2	1	U	ALL
Q01	a	Probability statements of likelihood	0.34	1	34	0.51	0.41	0.33	0.24	0.15	0.11	0.34
Q01	b	Probability statements of likelihood	0.92	1	92	0.99	0.98	0.97	0.91	0.76	0.41	0.92
Q01	c	Probability scale	0.85	1	85	0.98	0.95	0.89	0.78	0.65	0.31	0.85
Q01	d	Probability scale	0.66	1	66	0.93	0.83	0.66	0.45	0.30	0.10	0.66
Q02	a	Bar chart with probabilities	0.57	1	57	0.90	0.79	0.56	0.31	0.11	0.02	0.57
Q02	b	Bar charts - mode from	0.72	1	72	0.95	0.87	0.74	0.57	0.38	0.18	0.72
Q02	c	Bar charts - interpreting	0.36	1	36	0.60	0.48	0.34	0.21	0.09	0.02	0.36
Q03	a	Tally	1.65	2	83	1.89	1.84	1.71	1.55	1.20	0.51	1.65
Q03	b	Select an appropriate diagram	0.74	1	74	0.88	0.82	0.79	0.69	0.51	0.14	0.74
Q03	c	Calculate mode	0.56	1	56	0.87	0.73	0.54	0.36	0.19	0.04	0.56
Q03	d	Calculate median	0.57	2	28	1.24	0.72	0.39	0.22	0.14	0.05	0.57
Q03	e	Rationale for selecting types of average	0.52	2	26	0.79	0.64	0.52	0.35	0.17	0.05	0.52
Q04	a	Probability from a list	0.76	1	76	0.95	0.92	0.83	0.65	0.31	0.08	0.76
Q04	b	Compare data sets	1.77	4	44	3.10	2.44	1.67	0.73	0.24	0.03	1.77
Q04	c	Calculate using range	0.68	1	68	0.98	0.90	0.71	0.42	0.20	0.08	0.68
Q05		Planning data collection	1.02	6	17	2.39	1.36	0.67	0.26	0.10	0.01	1.02
Q06	a	Bar charts - completing	1.37	2	69	1.80	1.63	1.47	1.11	0.58	0.12	1.37
Q06	b	Bar charts - interpreting	0.60	1	60	0.85	0.74	0.60	0.45	0.23	0.08	0.60
Q06	c	Bar charts - interpreting	1.22	3	41	2.13	1.61	1.09	0.66	0.28	0.09	1.22
Q07	a	Definition of population	0.04	1	4	0.07	0.05	0.03	0.02	0.01	0.00	0.04
Q07	b	Definition of convenience sample	0.19	1	19	0.40	0.24	0.14	0.07	0.03	0.00	0.19
Q07	c	Disadvantages of sampling	0.19	1	19	0.45	0.25	0.13	0.06	0.02	0.00	0.19
Q07	d	Tabulation	0.60	2	30	1.01	0.76	0.57	0.36	0.15	0.03	0.60
Q07	e	Select an appropriate diagram	0.53	2	27	0.93	0.66	0.47	0.30	0.14	0.03	0.53
Q08	a	Select an appropriate diagram	0.13	1	13	0.32	0.16	0.07	0.04	0.02	0.00	0.13
Q08	b	Scatter diagrams - completing	0.95	2	48	1.65	1.30	0.90	0.44	0.13	0.01	0.95
Q08	c	Correlation	0.86	2	43	1.48	1.17	0.81	0.42	0.15	0.01	0.86
Q08	d	Determine line of best fit	0.93	2	47	1.65	1.29	0.85	0.40	0.15	0.04	0.93
Q08	e	Interpret Spearman's rank correlation coefficient	0.02	1	2	0.06	0.02	0.01	0.00	0.00	0.00	0.02
Q08	f	Interpret Spearman's rank correlation coefficient	0.25	2	13	0.51	0.35	0.21	0.08	0.03	0.00	0.25
Q08	g	Reliability and validity	0.20	2	10	0.65	0.25	0.06	0.02	0.01	0.00	0.20
Q09	a	Cleaning data	0.66	2	33	1.09	0.90	0.62	0.32	0.11	0.02	0.66
Q09	b	Select an appropriate diagram	0.17	1	17	0.40	0.22	0.11	0.04	0.01	0.00	0.17
Q10	a	Difference between primary and secondary data	0.49	1	49	0.80	0.67	0.47	0.25	0.10	0.02	0.49
Q10	b	Choropleth map - reading	0.66	1	66	0.92	0.84	0.69	0.47	0.25	0.07	0.66
Q10	c	Choropleth map - interpreting	0.79	2	40	1.32	1.12	0.76	0.35	0.12	0.01	0.79
Q11	a	Definitions for types of data	0.36	1	36	0.58	0.41	0.31	0.26	0.21	0.10	0.36
Q11	b	Box plots - reading	1.03	3	34	2.53	1.51	0.56	0.16	0.03	0.00	1.03
Q11	c	Box plots - comparing	0.74	5	15	2.17	0.98	0.26	0.05	0.01	0.00	0.74
Q11	d	Use stratification	0.10	3	3	0.37	0.09	0.02	0.01	0.00	0.00	0.10
Q12	a	Determine relative risks	0.36	3	12	0.90	0.50	0.21	0.06	0.02	0.00	0.36
Q12	b	Interpret relative risks	0.07	1	7	0.22	0.09	0.03	0.01	0.00	0.00	0.07
Q13	a	Median from cumulative frequency graph	0.27	1	27	0.55	0.31	0.20	0.14	0.09	0.02	0.27
Q13	b	Cumulative frequency graph - comparing	0.82	4	21	2.28	1.03	0.40	0.13	0.03	0.00	0.82
			27.29	80	34	47.04	34.83	24.37	15.38	8.41	2.79	27.29

Examiners Reports – F P2

Candidates responded well to the challenges within this paper, demonstrating understanding of a wide range of topics from the specification. It was pleasing to see students performed well on questions requiring standard techniques such as tally charts (Q03), comparing two means (Q04) and reading a choropleth map (Q10). They were generally confident at completing calculations and diagrams and demonstrated good statistical understanding when asked to interpret these. Like previous exam sessions candidates were still less familiar with certain topics e.g. stratified sampling, relative risk and cumulative frequency curves. It is worth mentioning here that centres must impress upon their students the importance of legible handwriting as a few responses were very difficult to read. Also, the use of phrase 'more accurate' is too often seen, and not usually awarded any marks. Students were generally able to attempt the whole paper within the time allowed, although some students made no attempt on some questions and many responses were left blank. Centres should also encourage students to show full working and set this out clearly so that partial credit can be awarded if a fully correct solution is not obtained.

Based on their performance on this paper, students should:

• Read each question fully and carefully before attempting to answer it. Check they have understood the requirement for comments in extended responses and try to give an answer for each section.

• Check graphs carefully for any part of the question asking you to complete it. Scatter diagrams will often (but not always) require additional points to be plotted. A bar chart with a blank space in the middle or at the end is likely to mean there is a requirement for them to complete it. Please encourage candidates to check where all marks are being awarded in a question.

• It is not enough to simply read and state values from a graph or boxplot, without forming some sort of comparison and conclusion. Candidates should check the wording of the question too. Often, these values are not required, and we are simply looking for visual comparisons to be made. Usually if a question requires these figures as part of your answer, the question will demand it.

• Candidates should practice interpreting answers to statistical calculations and diagrams in context. E.g. They should upgrade their responses to always comment on the comparison of IQR and the link to the data being less/more varied with context.

• Candidates should practice identifying the appropriateness of calculations, diagrams, and approaches in different settings.

• Candidates must comment on the appropriateness of a conclusion if the question has you to. Candidates should not assume we know what their decision is with regards to the appropriateness – we need to see a decision. 'No/yes because' is often the minimum required for these types of questions.

• Candidates should practice on making comparisons, particularly those in context.

• Candidates should appreciate that an item requiring an answer worth two marks will require two parts to their answer. A comment with a conclusion/decision. Or a correlation type with an interpretation.

• Candidates should revise how to calculate and interpret relative risk.

• Candidates should revise how to calculate the proportionate sizes for stratified sampling, and practice describing how to conduct this type of sampling method randomly.

• Candidates should revise cumulative frequency curves and reading values from it.

F P1 Q2c

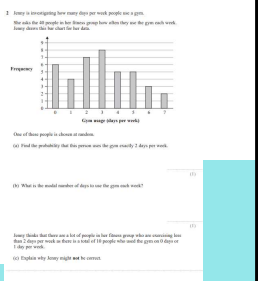
Jenny thinks that there are a lot of people in her fitness group who are exercising less than 2 days per week as there is a total of 10 people who used the gym on 0 days or 1 day per week.

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

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

(1)

(c)	B1 e.g. • 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	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. Allow 'the majority of people exercise more than two times a week' Condone sample of 40 is too small Do not allow • Gym use varies week by week • The question may have been misunderstood • 10 people is not a lot	(1)



B1 We will condone 'more than 2 days' and we will also forgive '30 more' instead of '30'.

F P1 Q2c

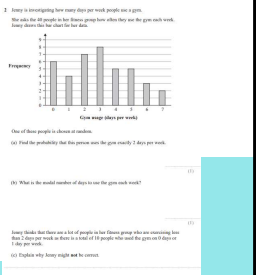
Jenny thinks that there are a lot of people in her fitness group who are exercising less than 2 days per week as there is a total of 10 people who used the gym on 0 days or 1 day per week.

(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)

(c)	B1 e.g. • 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	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. Allow 'the majority of people exercise more than two times a week' Condone sample of 40 is too small Do not allow • Gym use varies week by week • The question may have been misunderstood • 10 people is not a lot	(1)



B1 We will forgive the slip in units being written as hours instead of days.

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 its an average.

(2)

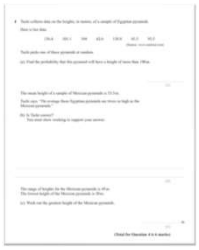
A1 2

(e)	B2 ft 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)	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) B0 for median with no attempt at reason	(2)

(e) B1 -- for median and an attempt at a reason.

(e) NOTE- we are not accepting the mode as the answer with any reason. Median only.

F P1 Q4b



The mean height of a sample of Mexican pyramids is 53.5 m.

Tachi says, "On average these Egyptian pyramids are twice as high as the Mexican pyramids."

(b) Is Tachi correct?
You must show working to support your answer.

mean of egyptian mountains = 100·28

53·5 × 2 = 107

Tachi is not corect because its not exactly 109 but was close.

(4)

M1 $\frac{136.4+101.1+104+62.6+138.8+65.5+93.5}{7}$ A1 100.2(7…) B1ft 107 or 50.1(3…) or 1.87… B1ft e.g. • yes as 107 and '100.2(7…)' are similar • no as 107 and '100.2(7..)' are not close enough in value • no as the mean for the Egyptian pyramids is '100.2(7…)' which is less than 107 • no as '100.27..' ÷ 53.5 = 1.87… < 2 therefore incorrect.	M1 for method to calculate the mean Condone 1 error in the numerator of the mean calculation A1 awrt 100.3 B1ft 107 or 50.1(3…) or 1.87… B1ft for evaluating the conclusion and a comparison made Follow through their 100.2(7…) provided this is from an attempt at calculating the mean.	(4)

M1A1 scored with 100.28 (awrt 100.3)

note: a correct answer for the mean implies correct method.

Award M1A1 if correct answer seen.

B1ft scored for 107

B1ft the comparison is implied and the conclusion is correc

F P1 Q5

B1 Comments on data collection: • 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 B1ft dep for appropriate comment on appropriateness on method of data collection consistent with the observations made.	B1 for a correct comment relating to the methods of data collection B1ft for comment on appropriateness of data collection consistent with their observations Dependent on previous B mark being awarded for data collection	(6)
B1 Comments on calculations: • 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. B1ft dep for appropriate comment on appropriateness on calculations consistent with the observations made.	B1 for a correct comment relating to the calculations B1ft for comment on appropriateness of calculations consistent with their observations Dependent on previous B mark being awarded for calculations	
B1 Comments on diagrams: • A bar chart would make it easier to compare 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. B1ft dep for appropriate comment on appropriateness on diagrams consistent with the observations made.	B1 for a correct comment relating to the diagrams B1ft for comment on appropriateness of diagrams consistent with their observations Dependent on previous B mark being awarded for diagrams	

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.

data collection is randomised which makes it accurate however she should ask for more than one month because it might have inaccurate results for one month

she should also change her diagrams so that it is more accessible to understand eg a scattergraph for average and box chart for range

Data Collection

B1 'she should ask for more than one month' no comment on appropriateness linked to this statement.

Calculations

B0 no correct comment seen

Diagrams

B0 incorrect comments given, we are not awarding suggestions for alternative diagrams

Note: had the candidate said that 'the bar chart is appropriate because it is more accessible for people to understand' this would be awarded B2. Allow a bar chart is easy to interpret / easy to understand / easy to compare DO NOT award a comment along the lines of 'a bar chart is easy to draw / takes less time / easy to read'

F P1 Q5

B1 Comments on data collection: • 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 B1ft dep for appropriate comment on appropriateness on method of data collection consistent with the observations made.	B1 for a correct comment relating to the methods of data collection B1ft for comment on appropriateness of data collection consistent with their observations Dependent on previous B mark being awarded for data collection	(6)
B1 Comments on calculations: • 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. B1ft dep for appropriate comment on appropriateness on calculations consistent with the observations made.	B1 for a correct comment relating to the calculations B1ft for comment on appropriateness of calculations consistent with their observations Dependent on previous B mark being awarded for calculations	
B1 Comments on diagrams: • A bar chart would make it easier to compare 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. B1ft dep for appropriate comment on appropriateness on diagrams consistent with the observations made.	B1 for a correct comment relating to the diagrams B1ft for comment on appropriateness of diagrams consistent with their observations Dependent on previous B mark being awarded for diagrams	

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

Data Collection

B1 'use secondary data, which will be quicker and easier' [note the candidate has 'They are appropriate, however this would take too long' - seems a contradictory statement to make - the candidate is not committing to appropriate or not appropriate]

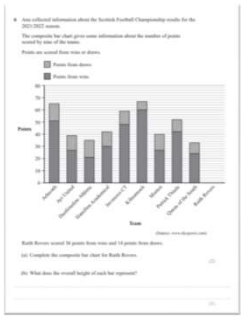
Calculations

B1 'she can (also) calculate a mean for each region ...' No comment on appropriateness linked to this specific statement Note: the sampling method does not specify which average, so if the mean or median is specifically mentioned in a response then it can be awarded B mark eg 'finding the mean of each region...' B1 eg ' 'finding the mean of each region is appropriate...' B2

Diagrams

B0 no valid comment on diagrams made

F P1 Q6b



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

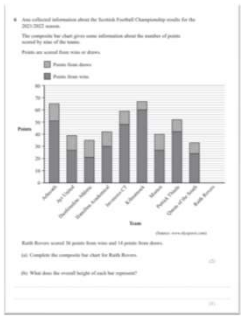
The amount of wins and draws.

(1)

	B1 e.g. the overall/total points scored/points from wins and draws		(1)

B0 This is not the same as the total points scored from wins and draws.

F P1 Q6b



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

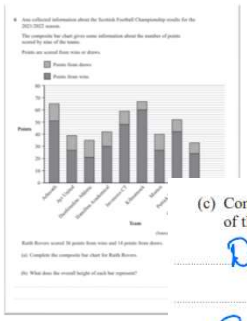
amount of goals scored

(1)

	B1 e.g. the overall/total points scored/points from wins and draws		(1)

B0 The graph is of points scored not goals

F P1 Q6c



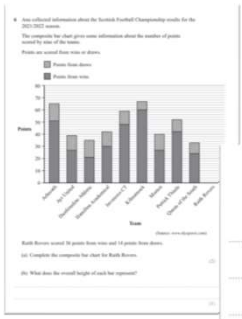
(c) Compare the points scored by Dunfermline Athletic with the points scored by Queen of the South.

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)

B1B1B1 Comparisons made of points scored, from wins, draws and in total. The numbers are not required.

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)

B0B0B0 'whereas' and 'whilst' are not enough on there own for the comparison. The candidate is just writing down values read from the graph. If the response had said 'Queen of the south scored 24 points from wins whereas Dunfermline Athleic only scored 21 points'' then this reads as a correct comparison

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. B0 for 'all people buying food'	
(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	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)

(a) B0 an acceptable answer must reference all / Every one of the visitors.

(b) B1 'such as standing at the theme park entrance and using the first 100 people' scores B for a description of how a convenience sample could be conducted. note: 'a convenience sample is something easy ' alone would score B0 We will award 'easy access (to people) but not that it is 'easy'

(c) B1 unlikely to be representative is enough to score this mark. Equivalent to saying 'not representative'

F P1 Q7 d/e

(d)	B2 for two comments from • (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 (B1 for one comment from the list)	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)	(2)

(e)	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)	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)	(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 datahes collecting

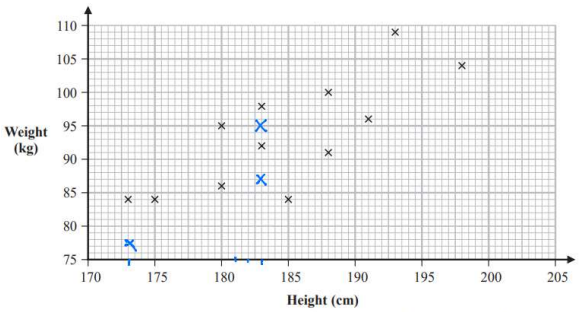
(2)

(d) B1 'shows a clear result' is equivalent to the suggested answer 'easy to analyse' in the mark scheme As mentioned before, we're not accepted answers that change the type of data that is collected (so opinions/descriptions/worded responses) If the candidate had written 'but does not leave room for added//other options' then this would be awarded the second mark.

(e) B1 for not appropriate, and there is an attempt at a reason

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)		(2)

(b) B2 All three points plotted correctly (they should all be on gridlines).

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)

B2 for e.g. • 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> (B1 for e.g. • 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	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)	(2)

B2 - for a correct comparison of the gradients, and concluding that Timur is correct.

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.

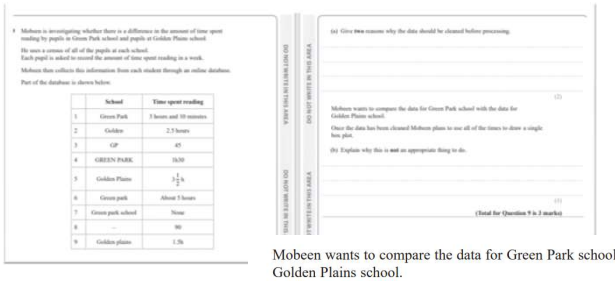
He is correct as England has a higher gradient of line of best fit meaning the weight increases faster as height increase

(2)

B2 for e.g. • Timur is correct as the gradient of the line of best fit of the England rugby squad backs is greater/steeper • Timur is not correct as the gradient of the lines of best fit is similar (B1 for e.g. • 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	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)	(2)

B0 Comparison of gradients is not correct - we will not accept 'higher' here. We will accept greater/steeper/larger/bigger. Note: 'He is correct' does not get awarded the B mark unless there is a correct comparison statement given to support this

F P1 Q9b



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.

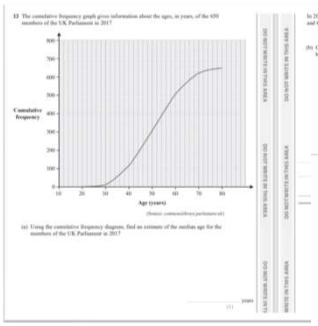
as he is comparing both teams so a single box won't help

(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)

B1 - this response is indicating that we need more than one box plot to compare

F P1 Q13b



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

500 - 215 = 285/600 = 0.475

47.5%

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

(4)

M1 '510' and '210'	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.	(4)
A1ft '300'	A1ft for their '300' (UB - LB) $280 \leq UB - LB \leq 320$	
A1ft for '46.1538…%'	A1ft for '46.1538…%' awrt '46.2% Allow for a percentage which is correct to 1 decimal place.	
B1ft for percentage of MPs (aged between 45 and 60) was greater in Germany	B1ft for comparing percentages Follow through their percentage for UK provided M1 scored.	

M1 A1 -- The UB, LB are correct and the difference between them has been found.

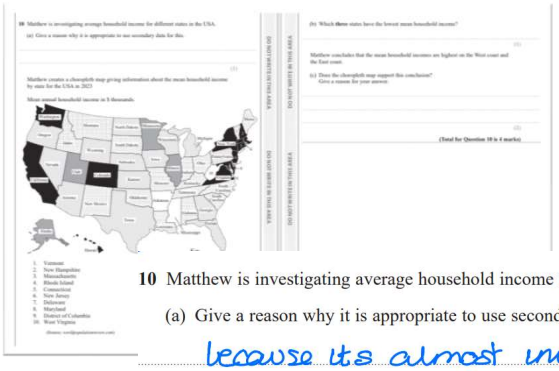
A0 percentage calculation is incorrect (600 used not 650)

B0ft Incorrect comparison made, percentages must be compared (this candidate has incorrectly assumed that a higher percentage means a larger number of members).

Pearson Edexcel

Foundation / Higher Crossover P2

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.

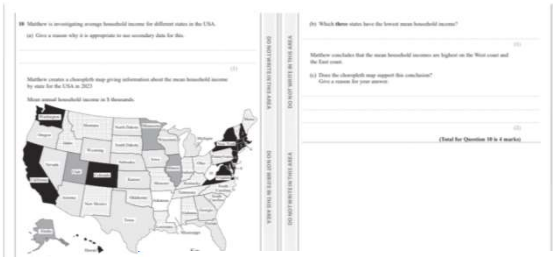
because its almost impossible to do by yourself

(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)

B1 -- we are allowing the use of the phrase 'impossible' to imply 'too difficult' provided it is clearly linked to primary data. In this case 'yourself' indicates the candidate means that primary would be 'impossible'

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 its 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)

B1 -- we are allowing the use of the phrase 'impossible' to imply 'too difficult' provided it is clearly linked to primary data. In this case 'yourself' indicates the candidate means that primary would be 'impossible'

F P1 Q11c

(c) Compare the three distributions of hand spans.
Give **three** comparisons and interpret **two** of your comparisons.

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

B1ft Correc... medians

B1ft Correct comparison of spread (IQR or range)

B1ft Correct comparison of skew

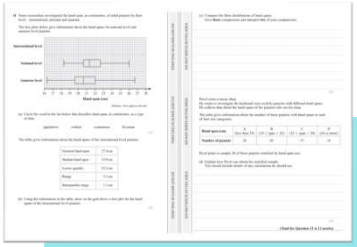
B1ft One correct interpretation

B1ft One further correct comparison of spread **or** interpretation

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

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



B0 - no comparison of medians

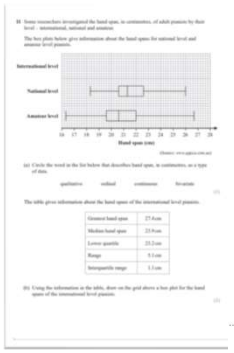
B0 - no comparison of skew

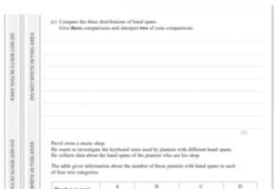
B1 - correct comparison of range

B0 - no interpretation of a comparison given

B0 - no further comparison of spread, IQR in this case. (nor is there a second interpretation) Note: a comparison of highest/greatest hand spans does not count as a further comparison. We are looking for an additional comparison on spread, in this case that would have been for a comparison made between at least two IQRs since the range has already been awarded.

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.

$24+65+57+14 = \frac{160}{20} = 8$

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

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

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

B1 Correct calculation of at least one sample size for a stratum.

B1 At least one correct sample size see for a stratum

B0

Higher P2

Skil

Q	Part	Skill tested	Mean score	Max score	Mean %	9	8	7	6	5	4	3	U	ALL
Q01	a	Difference between primary and secondary data	0.77	1	77	0.92	0.87	0.86	0.80	0.72	0.56	0.36	0.25	0.77
Q01	b	Choropleth map - reading	0.92	1	92	0.99	0.97	0.95	0.94	0.91	0.84	0.70	0.50	0.92
Q01	c	Choropleth map - interpreting	1.49	2	75	1.76	1.66	1.59	1.51	1.40	1.25	0.94	0.44	1.49
Q02	a	Definitions for types of data	0.71	1	71	0.97	0.92	0.83	0.72	0.57	0.43	0.31	0.32	0.71
Q02	b	Box plots - reading	2.69	3	90	2.92	2.93	2.88	2.83	2.67	2.11	1.16	0.50	2.69
Q02	c	Box plots - comparing	3.05	5	61	4.48	4.08	3.63	3.12	2.40	1.48	0.54	0.21	3.05
Q02	d	Use stratification	1.10	3	37	2.21	1.90	1.52	1.04	0.45	0.14	0.01	0.00	1.10
Q03	a	Determine relative risks	1.29	3	43	2.77	2.15	1.49	1.04	0.76	0.52	0.24	0.08	1.29
Q03	b	Interpret relative risks	0.21	1	21	0.40	0.29	0.25	0.20	0.17	0.09	0.02	0.02	0.21
Q04	a	Cumulative frequency- drawing	1.57	2	79	1.91	1.83	1.72	1.60	1.42	1.18	1.03	0.62	1.57
Q04	b	Cumulative frequency - using	1.25	2	63	1.89	1.77	1.56	1.28	0.92	0.46	0.20	0.10	1.25
Q05	a	Stem and leaf diagram	2.27	3	76	2.71	2.51	2.41	2.27	2.12	1.89	1.54	0.94	2.27
Q05	b	Calculate median	0.59	1	59	0.87	0.76	0.69	0.60	0.48	0.33	0.19	0.09	0.59
Q05	c	Calculate interquartile range	1.13	2	56	1.76	1.58	1.42	1.16	0.82	0.39	0.10	0.02	1.13
Q05	d	Compare data sets using appropriate measures	1.68	3	56	2.63	2.42	2.12	1.77	1.17	0.50	0.14	0.04	1.68
Q05	e	Comment on appropriateness	0.32	1	32	0.64	0.52	0.41	0.28	0.16	0.09	0.04	0.01	0.32
Q05	f	Limitations of data sources	0.40	1	40	0.72	0.60	0.48	0.38	0.28	0.13	0.02	0.01	0.40
Q06	a	Determine Spearman's rank correlation coefficient	3.23	6	54	5.23	4.77	4.24	3.30	2.11	0.88	0.30	0.27	3.23
Q06	b	Extraneous variables	0.47	2	24	1.04	0.76	0.56	0.43	0.27	0.11	0.03	0.00	0.47
Q06	c	Controlling extraneous variables	0.28	1	28	0.79	0.55	0.35	0.19	0.08	0.02	0.00	0.00	0.28
Q07	a	Use index numbers in context	1.34	2	67	1.91	1.87	1.71	1.39	0.94	0.54	0.25	0.08	1.34
Q07	b	Use index numbers in context	0.89	2	45	1.41	1.18	1.01	0.86	0.75	0.45	0.14	0.01	0.89
Q07	c	Calculate price index	0.50	2	25	1.80	1.27	0.58	0.19	0.04	0.01	0.01	0.00	0.50
Q07	d	Use appropriate measure of central tendency	0.39	1	39	0.73	0.61	0.48	0.34	0.25	0.18	0.09	0.03	0.39
Q07	e	Calculate geometric mean	0.98	3	33	2.04	1.78	1.38	0.84	0.40	0.11	0.03	0.00	0.98
Q08		Data Collection process	1.92	6	32	4.16	3.26	2.32	1.63	1.05	0.55	0.18	0.08	1.92
Q09	a	Interpret PMCC	0.95	2	48	1.79	1.53	1.23	0.89	0.51	0.30	0.18	0.06	0.95
Q09	b	Interpret regression equations	0.19	1	19	0.81	0.50	0.18	0.04	0.01	0.00	0.00	0.00	0.19
Q09	c	Use regression equations	0.27	3	9	1.10	0.57	0.25	0.14	0.07	0.03	0.01	0.01	0.27
Q09	d	Compare regression equations	0.21	2	11	1.13	0.50	0.17	0.04	0.01	0.00	0.00	0.00	0.21
Q09	e	Comment on appropriateness	0.37	2	19	0.94	0.64	0.44	0.28	0.20	0.09	0.04	0.00	0.37
Q10	a	Apply Petersen capture recapture formula	1.40	2	70	1.96	1.88	1.80	1.59	1.05	0.35	0.08	0.04	1.40
Q10	b	Reliability and replication	1.73	3	58	2.65	2.38	2.16	1.83	1.32	0.59	0.10	0.06	1.73
Q11	a	Venn diagrams	0.61	1	61	0.92	0.87	0.78	0.62	0.43	0.23	0.12	0.02	0.61
Q11	b	Formal notation for conditional probability	0.29	1	29	0.87	0.69	0.38	0.15	0.04	0.01	0.02	0.00	0.29
Q11	c	Formal notation for independent events	0.09	1	9	0.48	0.21	0.05	0.02	0.00	0.00	0.00	0.00	0.09
Q11	d	General addition law	0.70	2	35	1.82	1.42	0.94	0.47	0.20	0.08	0.02	0.00	0.70
			38.25	**80**	**48**	**64.13**	**55.00**	**45.82**	**36.78**	**27.16**	**16.92**	**9.14**	**4.81**	**38.25**

Examiners Reports – H P2

Most candidates have responded well to the challenges within this paper and demonstrated an understanding of a range of areas of the specification. It was pleasing to see candidates performed well on questions requiring standard techniques such as cumulative frequency diagrams (Q04), Spearman's rank correlation coefficient (Q06) and Peterson's capture recapture (Q10).

They were generally confident at completing calculations and diagrams, also demonstrated good statistical understanding when asked to interpret these. Candidates found questions requiring interpretation in context and evaluation of approaches or techniques slightly more challenging.

It is pleasing to report that questions requesting candidates to explain, assess or interpret have improved, compared to previous years.

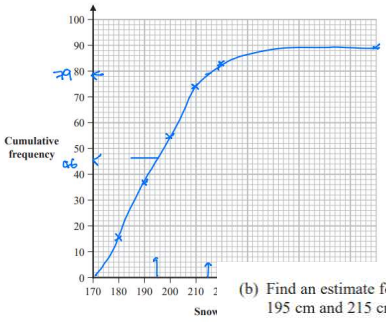
It is worth mentioning here that centres must impress upon their candidates the importance of legible handwriting as a few responses were found very difficult to read. Candidates were generally able to attempt the whole paper within the time allowed, although some candidates made no attempt on some questions.

Based on the performance of candidates on this paper, candidates should:

- Ensure that correct statistical language is used throughout when making comparisons, just stating figures, or using the word 'whereas' is not a comparison.
- Practice calculating relative risk and interpreting relative risk.
- Practise analysing plans for data collection and the subsequent diagrams and calculations together with giving statistical reasons or against the approaches suggested.
- Practice calculating geometric mean for a range of different contexts.
- Develop their understanding of independent events.
- Develop their understanding of regression equations and the interpretation of the gradient of these.

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.

79 − 46 = 33

33

(2)

		segments.	
	M1 for 80 − 46 A1ft 34	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)

(b) M1A1 an answer in the range 31-37.

Note: some candidates are working from the table in part (b), which would give answer of 32.5 which we be M1A1. If in doubt send item to review.

Note: we do not need to see lines on the graph from 195cm and 215cm to award these marks.

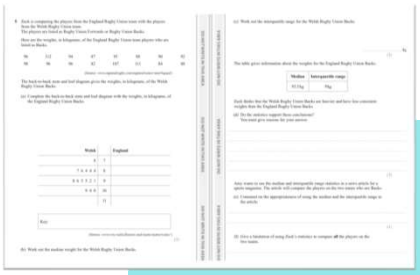
H P2 Q5a

Welsh		England
8	7	
7 6 4 4 4	8	88742
8 6 5 5 2 1	9	8864320
9 4 0	10	7
	11	20

Key: 8|8 is 88

(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)



B1 unordered stem and leaf diagram.

B0 unsuitable key.

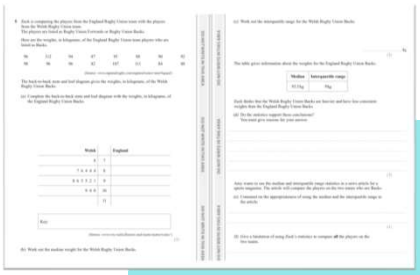
H P2 Q5a

Welsh		England
8	7	
7 6 4 4 4	8	24788
8 6 5 5 2 1	9	0234 6668
9 4 0	10	7
	11	12

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)



B2 correct stem and leaf diagram.

B0 unsuitable key. We allow the key as two parts but as this does not have the 2 lines either side of the stem and the correct numbers to match we would need to see the keys labelled Welsh and England.

H P2 Q5 e/f

(3)

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

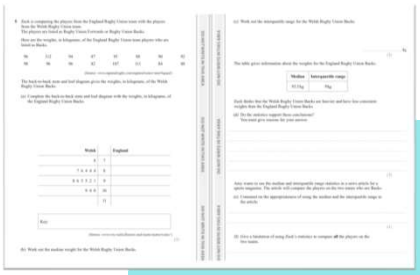
(1)

(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 effected 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)



(e) B0 as they have not said that it is 'not appropriate'. We need to see an assessment of the appropriateness with a reason

(f) B0 not one of the given reasons on the mark scheme. Do not allow reference to a disadvantage of the median or IQR

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)	B1 for e.g. • The same locations are tested without the drainage solution and then with the drainage solution B1 for e.g. • Different terrain. • Different surface materials e.g. tarmac. • Differences in forest cover area. • Different climate. SCB2 for testing for a year without the drainage solution and for a year with the drainage solution controls for seasons / time of year	B1 for identifying how the plan controls for extraneous variables. B1 for an example of an extraneous variable that is likely to be controlled for.	(2)
(c)	B1 for e.g. • Identify areas which have the same profiles e.g. same amount of rain, similar terrain.	B1 for describing a matched pairs approach where two areas are matched Allow for reference to two areas and matching of a feature. B0 for a description of testing the same area with drainage solution and then testing the same area after drainage solution.	(1)

(b) B1 measuring the same area B0 no example of a extraneous variable.

(c) B0 need to see identification of matching of a feature for two areas

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*)

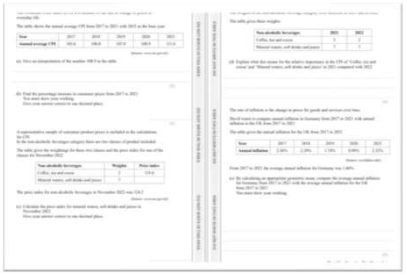
(a) Give an interpretation of the number 108.9 in the table.

8.9% increase

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

B2 for increase and 8.9%.

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.

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

$= 1.91\%$

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% B1depft on one previous mark scored.	(3)

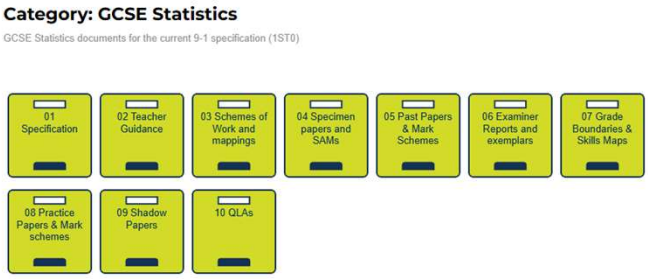
SCB1 for 1.91 or 1.91% B1 depft Correct comparison. This can be awarded as they have scored B1.

Support

Maths Emporium – Practice Papers & more

The **Maths Emporium** contains a rich source of resources for GCSE Statistics teachers, including:

- Practice papers including Statistical Enquiry practice and themed papers for the higher tier
- common question papers,
- past papers, mark schemes, examiner reports
- Shadow papers for some of the previous exam series
- *New* QLAs (standard and enhanced) for the Summer 2024 exam series
- Enhanced skills map for foundation and higher



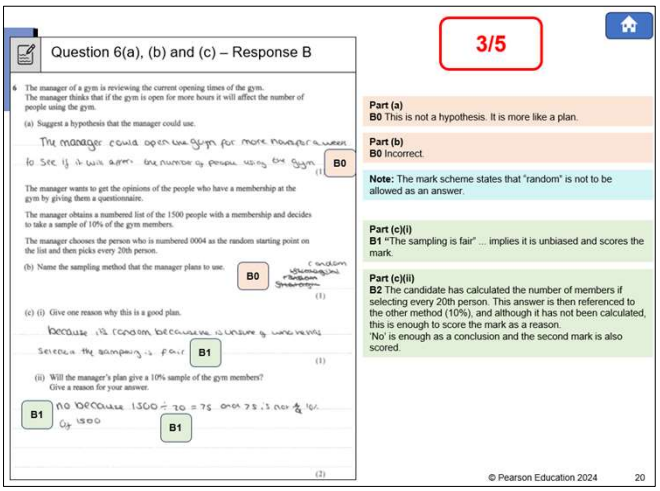
New Exemplars for GCSE Statistics

06 Examiner Reports and exemplars

For the summer 2024 exam series we have produced our exemplars in a new PowerPoint format and for the first time have created them for GCSE Statistics.

These are ready made slide decks with marking guidance on selected questions from the exam series. They use real candidate responses and have been annotated by the chair of examiners.

These have been designed for teachers to use for CPD but could also be used with students in the classroom.



The picture is an example slide from the foundation exemplar (P1 Q6)

Summer 2024 Post-Series Support and Upcoming Networks

GCSE Statistics

6th January 2025
- Reordered Papers
- New 'Aiming for' Practice Papers

Spring Term 2025
- New GCSE Maths and GCSE Statistics Crossover Question Papers

Still to be released from the Summer 2024 post-series support we have the reordered papers, *brand new* 'Aiming for' practice papers and *brand new* GCSE Maths and GCSE Statistics Crossover Papers.

Upcoming free online networks for GCSE Statistics:

- GCSE Statistics: Deep dive into lower performing topics
  Tuesday 10th December 2024 4 – 5pm
- GCSE Statistics: Deep dive into 'appropriateness'
  Thursday 6th January 2025 4 – 5pm
- GCSE Statistics: General revision guidance and extended answers
  Wednesday 26th March 2025 4 – 5pm

You can book these on the PD Academy here.

Contact us

Mel Muldowney
mel@justmaths.co.uk

Christian Seager
christian@justmaths.co.uk

JustMaths

Meet your Maths and Statistics Subject Advisor and Partner

Vicky Wood

Vicky works closely with the wider maths team to support teachers in UK centres in delivering Pearson Edexcel qualifications in Mathematics and Statistics teachingmaths@pearson.com
Sign up for Vicky's monthly updates

Nicola Woodford-Smith

Nicola works as the maths Subject Partner in the maths team at Pearson Edexcel. She helps to create resources and delivers CPD to support you and your team through the lifecycle of our qualifications.
Follow @miss_mathsgeek (on X) for updates and information

VW

NEW Level 2 Extended Maths Certificate

Giving students the **opportunity to challenge** themselves at Key Stage 4 and build the **perfect foundation** for further study.

Scan the QR code or use this link to **register your interest** and to find out more.

You can also find the recordings of our launch event and Getting Ready to Teach event on the Maths Emporium.

NEW Podcast: The Right Angle

The Right Angle invites topical discussions, debates and insights from a range of thought leaders, award-winning maths educators and facilitators. Our subject partner hosts, Mark Heslop and Nicola Woodford-Smith lead conversations on themes such as the evolution of technology to support learning, student engagement and diversity and inclusion across the education of mathematics. Listen and subscribe for FREE on Apple Podcasts, Spotify and on Soundcloud.

Pearson

Find out more

For more professional development courses please see Pearson's [Professional Development Academy]

Your Feedback Matters

Following this event, you will receive an invitation to share your thoughts about the session. Your feedback is invaluable to us, as it helps us tailor our professional development materials to better meet your needs. Please don't hesitate to let us know what you'd like to see more of and what areas you think could be improved.


Pearson