GCSE Statistics

Summer 2023 Marking Guidance

Melanie Muldowney
Christian Seager

Pearson
Edexcel

Agenda

- General marking guidance
- Specific questions
- Examiners reports


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

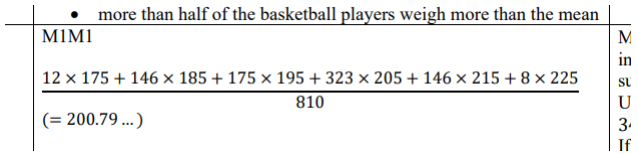
Probability answers must be given as a fraction, percentage or decimal. If a candidate gives a decimal equivalent to a probability, this should be written to at least 2 decimal places (unless tenths).
Incorrect notation should lose the accuracy marks, but be awarded any implied method marks.
If a probability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

Paper 1

Pearson
Edexcel

Foundation Paper 1

Q	Part	Skill tested	Score	Max score	Percent	ALL	1	2	3	4	5	6	7	8	9	ALL	U
Q01	a	Reading tables	0.95	1	95	0.95	0.89	0.93	0.96	0.98	0.98					0.95	0.72
Q01	b	Reading tables	0.95	1	95	0.95	0.88	0.94	0.97	0.97	0.98					0.95	0.63
Q01	c	Using tables	1.48	2	74	1.48	0.67	1.18	1.62	1.78	1.87					1.48	0.31
Q01	d	Comparing averages from tables	0.73	1	73	0.73	0.45	0.61	0.74	0.85	0.93					0.73	0.24
Q01	e	Justify the appropriate format and produce accurate visualisation of data	0.25	1	25	0.25	0.07	0.13	0.23	0.32	0.47					0.25	0.01
Q02	a	Probability Scale	0.88	1	88	0.88	0.7	0.85	0.92	0.95	0.98					0.88	0.36
Q02	b	Describing likelihood	0.94	1	94	0.94	0.82	0.96	0.97	0.98	0.99					0.94	0.46
Q02	c	Relative frequency	1.04	2	52	1.04	0.37	0.62	0.97	1.41	1.72					1.04	0.1
Q02	d	Compare probability of outcomes	1.08	2	54	1.08	0.46	0.79	1.1	1.35	1.55					1.08	0.11
Q03		Misleading graphs/diagrams	0.96	2	48	0.96	0.4	0.72	0.97	1.21	1.4					0.96	0.1
Q04	a	Data collection	0.79	2	40	0.79	0.28	0.56	0.76	0.99	1.3					0.79	0.06
Q04	b ii	Population, sample frame and sample	0.14	1	14	0.14	0.04	0.06	0.11	0.18	0.32					0.14	0.01
Q04	b ii	Population, sample frame and sample	0.01	1	1	0.01				0.01	0.03					0.01	
Q04	c	Strategies to mitigate issues during the statistical enquiry cycle	0.2	2	10	0.2	0.02	0.06	0.12	0.26	0.54					0.2	0.01
Q04	d	Key features when planning data collection	1.3	6	22	1.3	0.16	0.49	1.03	1.85	2.93					1.3	0.03
Q05	a	Vocabulary of correlation	1.06	2	53	1.06	0.28	0.67	1.1	1.39	1.64					1.06	0.03
Q05	b	Determine line of best fit	0.66	1	66	0.66	0.34	0.54	0.71	0.78	0.87					0.66	0.06
Q05	c	Using a line of best fit	0.43	1	43	0.43	0.11	0.27	0.44	0.57	0.69					0.43	
Q05	d	Justify appropriateness of using a line of best fit to make predictions	0.32	2	16	0.32	0.04	0.12	0.26	0.43	0.75					0.32	0.01
Q06	a	Terms used to describe different types of data	0.25	1	25	0.25	0.11	0.15	0.2	0.32	0.5					0.25	0.04
Q06	b	Calculate mode	0.67	1	67	0.67	0.31	0.55	0.7	0.82	0.9					0.67	0.1
Q06	c	Calculate median	0.8	1	80	0.8	0.5	0.73	0.85	0.92	0.94					0.8	0.16
Q06	d	Justify the rationale for selecting appropriate types of average	0.32	1	32	0.32	0.08	0.21	0.35	0.42	0.45					0.32	0.01
Q06	e	Calculate interquartile range (IQR)	0.59	2	30	0.59	0.16	0.28	0.47	0.78	1.32					0.59	0.02
Q06	f	Use median and interquartile range (IQR) to compare data	0.43	4	11	0.43	0.01	0.04	0.2	0.57	1.46					0.43	
Q06	g i	Comment on outliers with reference to the original data	0.13	1	13	0.13	0.01	0.02	0.08	0.2	0.33					0.13	
Q06	g ii	Comment on outliers with reference to the original data	0.18	1	18	0.18	0.03	0.06	0.16	0.27	0.36					0.18	
Q07		Use stratification and know when this is appropriate before sampling	0.09	5	2	0.09		0.01	0.03	0.09	0.39					0.09	
Q08	a	Sampling techniques to avoid bias	0.51	2	26	0.51	0.18	0.32	0.5	0.63	0.9					0.51	0.04
Q08	b	Sampling techniques to avoid bias	0.04	1	4	0.04		0.01	0.02	0.06	0.11					0.04	
Q08	c	Sampling techniques to avoid bias	0.14	2	7	0.14	0.01	0.03	0.07	0.16	0.48					0.14	
Q08	d	Use summary statistics and sample data to estimate or predict population characteristics	0.04	2	2	0.04		0.01	0.02	0.05	0.16					0.04	
Q09	a	Terms used to describe different types of data	0.33	1	33	0.33	0.15	0.23	0.29	0.4	0.56					0.33	0.06
Q09	b	Histograms (equal class width) - completing a table	1.07	2	54	1.07	0.17	0.52	1.07	1.55	1.77					1.07	0.02
Q09	c	Histograms (equal class width) - completing a histogram	0.89	2	45	0.89	0.09	0.34	0.86	1.33	1.61					0.89	0.02
Q09	d	Interpret a distribution of data in term of skewness	0	1	0						0.01						
Q09	e i	Calculate an estimate of the mean	0.59	3	20	0.59	0.04	0.14	0.38	0.84	1.61					0.59	
Q09	e ii	Compare data sets using appropriate measure of central tendency	0.43	1	43	0.43	0.05	0.17	0.38	0.62	0.83					0.43	0.01
Q10	a	Hypothesis testing	0.13	1	13	0.13		0.01	0.07	0.18	0.39					0.13	
Q10	b	Interpret seasonal and cyclic trends in context	0.38	2	19	0.38	0.01	0.07	0.26	0.58	0.98					0.38	
Q10	c	Comment of appropriateness of using moving averages	0.12	1	12	0.12	0.01	0.02	0.09	0.18	0.3					0.12	
Q10	d	Compare seasonal and cyclic trends in context	0.09	1	9	0.09		0.01	0.03	0.12	0.31					0.09	
Q10	e	Time series - Plotting	0.46	3	15	0.46	0.03	0.1	0.31	0.68	1.21					0.46	
Q10	f	Interpret seasonal and cyclic trends in context	0.37	2	19	0.37	0.05	0.14	0.29	0.5	0.87					0.37	
Q11	a	Sample space diagrams - Completing	1.33	2	67	1.33	0.54	0.93	1.39	1.7	1.88					1.33	0.14
Q11	b	Sample space diagrams - probability from	0.77	2	39	0.77	0.1	0.3	0.68	1.12	1.53					0.77	0.02
			25.32	80		25.32	9.62	15.9	23.73	32.35	43.1	0	0	0	0	25.32	3.89

The table shows the average heights, to the nearest cm, of Adult Males and Adult Females who were born in the year 1996 in some countries of the world.

Country	Average height (cm)		Difference
	Adult Males	Adult Females	
Philippines	163	150	13
Latvia	181	170	11
Italy	178	165	13
Zimbabwe	169	158	11
Australia	180	166	14

(Source: *Ourworldindata.org*)

(a) Write down the average height of Adult Males in Italy.

178 ✓ cm

(1)

(b) Write down the country in the table in which the Adult Females have the greatest average height.

170 ✗

(1)

Afzal thinks that the country in the table with the greatest difference in average heights between Adult Males and Adult Females is Italy.

(c) Is Afzal correct?
Give a reason for your answer.

No, Australia is bigger difference ✓
14 > 13 ✓

(2)

(d) Using the table, compare the average height of Adult Males in Australia, with the average height of Adult Males in Zimbabwe.

Australian males are taller than males in Zimbabwe ✓

(1)

Afzal suggests drawing a time series graph to represent the data in the table.

(e) Explain whether or not this is an appropriate graph to use.

A bar chart would be better ✗

(1)

(Total for Question 1 is 6 marks)

1(a)	B1 for 178		**(1)**
(b)	~~B1 for Latvia~~		**(1)**
(c)	B2 for the difference in Italy is 13 **and** in Australia(Philippines) it is 14 (13), so no/Afzal is not correct (B1 for any correctly identified difference)	Award B2 for the differences compared between two relevant countries and a correct conclusion. Must see 2 relevant differences next to table or in written answer. Allow B2 for No, Australia has a 1 cm **greater difference**. Figure must be correct for this. SC B1 for the difference in Italy and the Philippines is the same. For B2 or B1 Accept correct differences seen by table.	**(2)**
(d)	B1 for (Adult Males are) taller in Australia than (Adult Males) in Zimbabwe.	B1 for Adult Males are 11cm taller in Australia Ignore incorrect figures if there is an indication that Australian males are taller oe. Allow B1 for there is an 11cm difference in heights. Figure must be correct for this. For taller accept higher/greater/larger/biggest/(average) height	**(1)**
(e)	B1 For one of e.g. • Not appropriate as no time was given/recorded. • Not appropriate because only have data for one year	Do not accept: • given data about heights not time Condone 'data does not use time'	**(1)**

Foundation Paper 1 Q1

Student A

The table shows the average heights, to the nearest cm, of Adult Males and Adult Females who were born in the year 1996 in some countries of the world.

Country	Average height (cm)	
	Adult Males	Adult Females
Philippines	163	150
Latvia	181	170
Italy	178	165
Zimbabwe	169	158
Australia	180	166

(Source: *Ourworldindata.org*)

(a) Write down the average height of Adult Males in Italy.

178 ✓ cm

(1)

(b) Write down the country in the table in which the Adult Females have the greatest average height.

Latvia ✓

(1)

Afzal thinks that the country in the table with the greatest difference in average heights between Adult Males and Adult Females is Italy.

(c) Is Afzal correct?
Give a reason for your answer.

No, he is wrong Australias difference is greater ✗

(2)

(d) Using the table, compare the average height of Adult Males in Australia, with the average height of Adult Males in Zimbabwe.

180 - 169 = 11 ✗

(1)

Afzal suggests drawing a time series graph to represent the data in the table.

(e) Explain whether or not this is an appropriate graph to use.

No it is not appropriate. ✗

(1)

(Total for Question 1 is 6 marks)

1(a)	B1 for 178		**(1)**
(b)	B1 for Latvia		**(1)**
(c)	B2 for the difference in Italy is 13 **and** in Australia(Philippines) it is 14 (13), so no/Afzal is not correct (B1 for any correctly identified difference)	Award B2 for the differences compared between two relevant countries and a correct conclusion. Must see 2 relevant differences next to table or in written answer. Allow B2 for No, Australia has a 1 cm **greater difference**. Figure must be correct for this. SC B1 for the difference in Italy and the Philippines is the same. For B2 or B1 Accept correct differences seen by table.	**(2)**
(d)	B1 for (Adult Males are) taller in Australia than (Adult Males) in Zimbabwe.	B1 for Adult Males are 11cm taller in Australia Ignore incorrect figures if there is an indication that Australian males are taller oe. Allow B1 for there is an 11cm difference in heights. Figure must be correct for this. For taller accept higher/greater/larger/biggest/(average) height	**(1)**
(e)	B1 For one of e.g. • Not appropriate as no time was given/recorded. • Not appropriate because only have data for one year	Do not accept: • given data about heights not time Condone 'data does not use time'	**(1)**

Foundation Paper 1 Q1

Student B

Foundation Paper 1 Q1

WAGOLL

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 The table shows the average heights, to the nearest cm, of Adult Males and Adult Females who were born in the year 1996 in some countries of the world.

Country	Average height (cm)		Difference
	Adult Males	Adult Females	
Philippines	163	150	13
Latvia	181	170	11
Italy	178	165	13
Zimbabwe	169	158	11
Australia	180	166	14

(Source: *Ourworldindata.org*)

(a) Write down the average height of Adult Males in Italy.

178 cm

(1)

(b) Write down the country in the table in which the Adult Females have the greatest average height.

Latvia

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Afzal thinks that the country in the table with the greatest difference in average heights between Adult Males and Adult Females is Italy.

(c) Is Afzal correct?
Give a reason for your answer.

No, the difference for Italy is 13 but in Australia its 14, so he is not correct.

(2)

(d) Using the table, compare the average height of Adult Males in Australia (180), with the average height of Adult Males in Zimbabwe (169).

Australian adult males are on average taller than adult males in Zimbabwe.

(1)

Afzal suggests drawing a time series graph to represent the data in the table.

(e) Explain whether or not this is an appropriate graph to use.

No, it is not appropriate as no time has been given, the data is for one year only

(1)

(Total for Question 1 is 6 marks)

A town council is proposing to build a new leisure centre.
Michelle is going to carry out a survey to find out what all the people in the town think of the proposal.

Michelle thinks that she should take a sample rather than a census.

(a) Give **two** reasons why Michelle might think this.

- cheap ✓
- easy ✓

(2)

Michelle plans to use the electoral register as the sampling frame.

(b) (i) Explain what you understand by the term sampling frame.

A frame used for sampling. ✗

(1)

(ii) Give one problem Michelle may have using the electoral register as the sampling frame.

inaccurate ✗

(1)

Michelle intends to conduct a pilot study.

(c) Give **two** reasons why it is a good idea to conduct a pilot study.

using a sample is quick and less data is collected too ✗ ✗

(2)

uestion umber	Answer	Additional guidance	Mark
4 (a)	B1 B1 for two from • easier / difficult to ensure whole population is used • cheaper • quicker • less data to handle/fewer people to ask	Allow converse statement if census mentioned. Do not accept • not everyone will want to take part in the survey • she can choose who she wants to survey • reference to being able to explain the questions	(2)
(b)(i)	B1 A list of **all** members of the population/**all** people (who live) in the town	Do not accept 'whole group'	(1)
(b)(ii)	B1 for one from • children will not be included • only those registered to vote are on the list • difficulty in gaining access to electoral register/electoral register is confidential • the information the council has may be out of date/missing data/inaccurate	Must be a relevant comment about the electoral register. Do not accept one word answers such as 'inaccurate/unreliable/bias' Condone Electoral register is unreliable	(1)
(c)	B1 B1 for any two from: • The pilot study identifies problems • Checks that questions work as intended • Gives an idea of what results may be (for suitability for analysis) • Tests questions are clear/understood • Gives an idea of response rate • Checks questions are inoffensive		(2)

Foundation Paper 1
Q4 (a to c)

Student A

A town council is proposing to build a new leisure centre.
Michelle is going to carry out a survey to find out what all the people in the town think of the proposal.

Michelle thinks that she should take a sample rather than a census.

(a) Give **two** reasons why Michelle might think this.

- it would be difficult to ask the whole population ✓

✗

(2)

Michelle plans to use the electoral register as the sampling frame.

(b) (i) Explain what you understand by the term sampling frame.

it is a list of all the members of the population. ✓

(1)

(ii) Give one problem Michelle may have using the electoral register as the sampling frame.

getting hold of the electoral register maybe difficult. ✓

(1)

Michelle intends to conduct a pilot study.

(c) Give **two** reasons why it is a good idea to conduct a pilot study.

- gives an idea about response rate ✓
- tests the survey to check it works as expected ✓

(2)

Question number	Answer	Additional guidance	Mark
4 (a)	B1 B1 for two from • easier / difficult to ensure whole population is used • cheaper • quicker • less data to handle/fewer people to ask	Allow converse statement if census mentioned. Do not accept • not everyone will want to take part in the survey • she can choose who she wants to survey • reference to being able to explain the questions	(2)
(b)(i)	B1 A list of **all** members of the population/**all** people (who live) in the town	Do not accept 'whole group'	(1)
(b)(ii)	B1 for one from • children will not be included • only those registered to vote are on the list • difficulty in gaining access to electoral register/electoral register is confidential • the information the council has may be out of date/missing data/inaccurate	Must be a relevant comment about the electoral register. Do not accept one word answers such as 'inaccurate/unreliable/bias' Condone Electoral register is unreliable	(1)
(c)	B1 B1 for any two from: • The pilot study identifies problems • Checks that questions work as intended • Gives an idea of what results may be (for suitability for analysis) • Tests questions are clear/understood • Gives an idea of response rate • Checks questions are inoffensive		(2)

Foundation Paper 1
Q4 (a to c)

Student B

Michelle is writing a plan for her investigation into people's views on the leisure centre proposal.

(d) Write down what Michelle should include in her plan.
You should include each of the following

- a sampling method
- a question she could ask in her questionnaire
- a statistical diagram she could use to show the results of the survey.

Explain why each of the things you have written down is appropriate.

She should use random sampling to select who to send her survey to and then use lots of diagrams to display her data : bar charts, pie charts, pictograms

Sampling method B1 for appropriate sampling method identified, one of: • Random sample • Stratified sample • Systematic sample B1 Reason – any from • every resident has an equal chance of being selected (random) • sample selected will be unbiased (random/stratified) • (if n is large) sample is representative of the population (random/stratified/systematic) Question B1 for an appropriately designed (or described design) question. depB1 Reason e.g. for explaining that their question/question design achieves one or more of: • finds people views on the leisure centre proposal/information we want, • is non-leading • is closed/has response options Statistical diagram B1 For identifying a suitable statistical diagram e.g. Bar chart/line chart/pictogram/pie chart depB1 For reason: e.g. … shows frequencies/allows comparison	Accept a correct description of conducting one of these appropriate sampling methods e.g. using a numbered list and random number generator/selecting names from a hat (random sample) e.g. a description of the strata and selecting randomly within these (stratified sample) Do not accept even chance. Appropriate must be: • a non-leading closed question • with written (or implied) options for response. **If more than one question is suggested mark all and award highest score. B1B1 can only be awarded for one suggested question, not across multiple.** e.g B1 Pie chart because… depB1 it shows proportions/percentages Do not accept Tally chart **If more than one diagram is suggested mark all and award highest score. B1B1 can only be awarded for one suggested diagram, not across multiple.**	**(6)**

Foundation Paper 1
Q4 (d)

Student A

Foundation Paper 1 Q4

WAGOLL

4 A town council is proposing to build a new leisure centre.
Michelle is going to carry out a survey to find out what all the people in the town think of the proposal.

Michelle thinks that she should take a sample rather than a census.

(a) Give **two** reasons why Michelle might think this.

Two from

eg [1] it will be cheaper [2] there will be less data to handle

[3] it will be quicker

[4] it would be difficult to ask the whole population

(2)

Michelle plans to use the electoral register as the sampling frame.

(b) (i) Explain what you understand by the term sampling frame.

it is a list of all the members of the population (in this case a list of all the people who live in the town)

(1)

(ii) Give one problem Michelle may have using the electoral register as the sampling frame.

e.g. only people registered to vote will be on the list.

(1)

Michelle intends to conduct a pilot study.

(c) Give **two** reasons why it is a good idea to conduct a pilot study.

eg Two from:-

[1] any problems can be identified

[2] it will test that the questions are clear and understood

[3] gives an indication of the response rate

(2)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Michelle is writing a plan for her investigation into people's views on the leisure centre proposal.

(d) Write down what Michelle should include in her plan.
You should include each of the following

allowed methods: (random, stratified or systematic)

- a sampling method
- a question she could ask in her questionnaire
- a statistical diagram she could use to show the results of the survey.

(eg bar chart, pictogram, line graph, pie chart)

Explain why each of the things you have written down is appropriate.

A random sample, ie selecting names from a hat should be used as every resident would have an equally likely chance of being selected.

A possible question would be:-

The local council is proposing to build a new leisure centre. Do you...

☐ strongly support? ☐ support? ☐ have no opinion ☐ do not support ☐ not relevant/ prefer not to say

She could draw a pie chart as it would allow her to show proportions and allow comparisons.

(6)

(Total for Question 4 is 12 marks)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

5 Jim is investigating the relationship between air temperature and altitude.

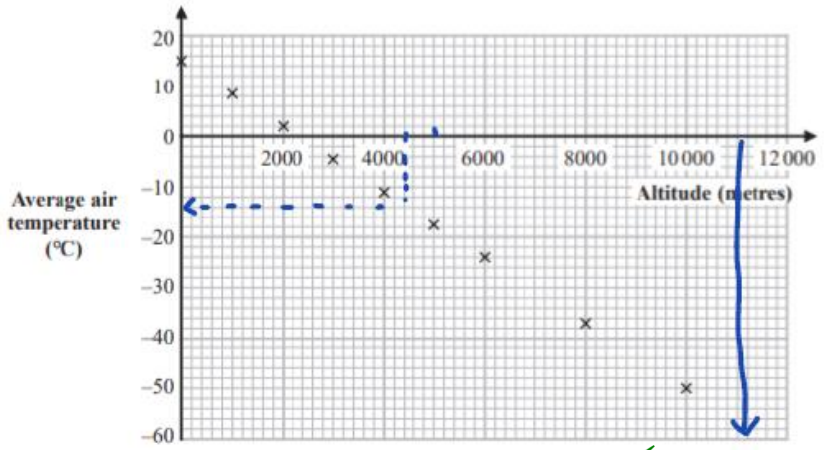
He has found data on the internet showing the average air temperature at different altitudes on one particular day.

The table shows the results he collected.

Altitude (metres)	0	1000	2000	3000	4000	5000	6000	8000	10000
Average air temperature (°C)	15.0	8.5	2.0	−4.5	−11.0	−17.5	−24.0	−36.9	−49.9

(Source: adapted from www.engineeringtoolbox.com)

He uses this information to draw the scatter diagram.



(a) Describe and interpret the correlation shown in the scatter diagram.

There is a negative correlation, which means as the altitude increases the air temp. decreases.

(2)

(b) Draw a line of best fit on the scatter diagram.

(1)

(c) Use the scatter diagram and your line of best fit to predict the average air temperature at an altitude of 4400 metres.

−14

(1)

Jim wants to predict the average air temperature at an altitude of 11 000 metres.

(d) Discuss whether or not it is appropriate to use the line of best fit to make his prediction.

The points only go as far as 10000

(2)

(Total for Question 5 is 6 marks)

5(a)	B2 for negative (correlation) so as the altitude increases, temperature decreases (B1 for either **negative** (correlation) **or** an explanation that the higher the altitude, the lower the temperature)	For B2 or B1 Accept converse when describing the correlation eg. (negative), as the temperature increases the altitude decreases. Ignore reference to strength of correlation.	**(2)**
(b)	B1 for the line of best fit plotted correctly	Should go through all the points and be drawn with a ruler.	**(1)**
(c)	B1 for −14	B0 if there is no line of best fit. B1ft their line of best fit Allow tolerance on their '−14' of half a small square. For follow through, the gradient of line of best fit must be negative.	**(1)**
(d)	B2 for extrapolating/outside of range for altitude/data only goes up to 10000(m) is not reliable so it is therefore not appropriate (B1 for reference to extrapolation/outside of range for altitude/data only goes up to 10000(m) with no or incorrect conclusion) or B2 for perfect correlation/strong correlation/all points lie on a straight line AND 11000 is close to the range of values for altitude therefore it is appropriate. (B1 for reference to strong correlation/perfect correlation/points lie close to the line of best fit with no or incorrect conclusion)	Allow for beyond/outside of the data plotted or beyond/outside within the table.	**(2)**

Foundation Paper 1 Q5 Student A

Jim is investigating the relationship between air temperature and altitude.

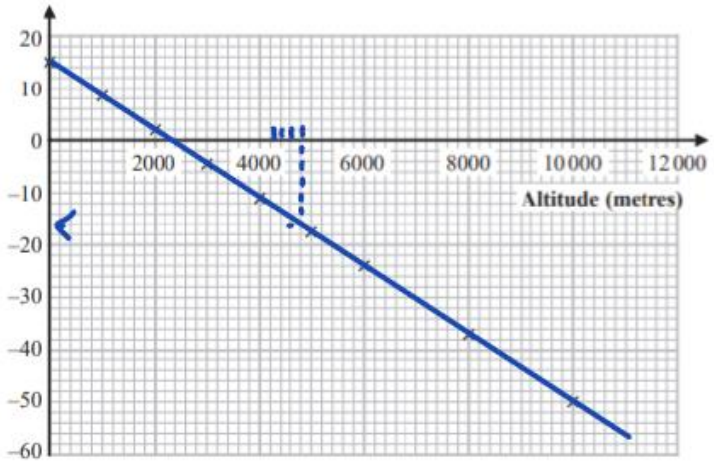
He has found data on the internet showing the average air temperature at different altitudes on one particular day.

The table shows the results he collected.

Altitude (metres)	0	1000	2000	3000	4000	5000	6000	8000	10000
Average air temperature (°C)	15.0	8.5	2.0	−4.5	−11.0	−17.5	−24.0	−36.9	−49.9

(Source: adapted from www.engineeringtoolbox.com)

He uses this information to draw the scatter diagram.



(a) Describe and interpret the correlation shown in the scatter diagram.

its negative ✓ ✗

(2)

(b) Draw a line of best fit on the scatter diagram. ✓

(1)

(c) Use the scatter diagram and your line of best fit to predict the average air temperature at an altitude of 4400 metres.

✗ -16

(1)

Jim wants to predict the average air temperature at an altitude of 11000 metres.

(d) Discuss whether or not it is appropriate to use the line of best fit to make his prediction.

In this case it is appropriate as all the lines sit on the line of best fit. ✓ ✓

(2)

(Total for Question 5 is 6 marks)

5(a)	B2 for negative (correlation) so as the altitude increases, temperature decreases (B1 for either **negative** (correlation) **or** an explanation that the higher the altitude, the lower the temperature)	For B2 or B1 Accept converse when describing the correlation eg. (negative), as the temperature increases the altitude decreases. Ignore reference to strength of correlation.	**(2)**
(b)	B1 for the line of best fit plotted correctly	Should go through all the points and be drawn with a ruler.	**(1)**
(c)	B1 for −14	B0 if there is no line of best fit. B1ft their line of best fit Allow tolerance on their '−14' of half a small square. For follow through, the gradient of line of best fit must be negative.	**(1)**
(d)	B2 for extrapolating/outside of range for altitude/data only goes up to 10000(m) is not reliable so it is therefore not appropriate (B1 for reference to extrapolation/outside of range for altitude/data only goes up to 10000(m) with no or incorrect conclusion) or B2 for perfect correlation/strong correlation/all points lie on a straight line AND 11000 is close to the range of values for altitude therefore it is appropriate. (B1 for reference to strong correlation/perfect correlation/points lie close to the line of best fit with no or incorrect conclusion)	Allow for beyond/outside of the data plotted or beyond/outside within the table.	**(2)**

Foundation Paper 1 Q5 Student B

Foundation Paper 1 Q5

5 Jim is investigating the relationship between air temperature and altitude.

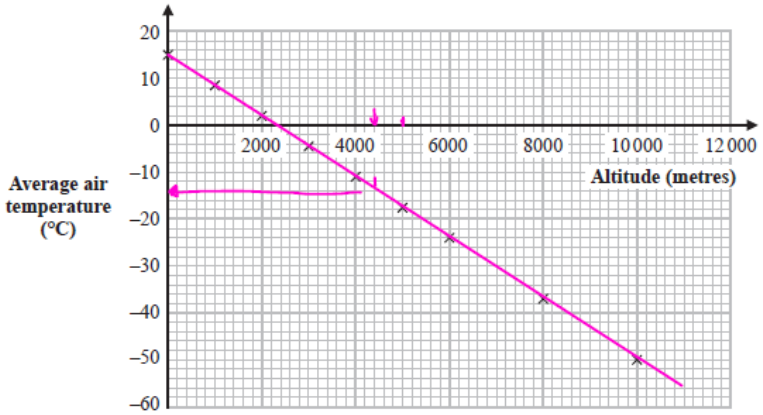
He has found data on the internet showing the average air temperature at different altitudes on one particular day.

The table shows the results he collected.

Altitude (metres)	0	1000	2000	3000	4000	5000	6000	8000	10 000
Average air temperature (°C)	15.0	8.5	2.0	–4.5	–11.0	–17.5	–24.0	–36.9	–49.9

(Source: adapted from *www.engineeringtoolbox.com*)

He uses this information to draw the scatter diagram.



(a) Describe and interpret the correlation shown in the scatter diagram.

Negative correlation, so as the altitude decreases the air temperature also decreases.

(2)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

(b) Draw a line of best fit on the scatter diagram.

(1)

(c) Use the scatter diagram and your line of best fit to predict the average air temperature at an altitude of 4400 metres.

-14

(1)

Jim wants to predict the average air temperature at an altitude of 11 000 metres.

(d) Discuss whether or not it is appropriate to use the line of best fit to make his prediction.

This would not be appropriate because the given data only goes to 10,000m and 11,000 would be outside this so would be unreliable.

(2)

(Total for Question 5 is 6 marks)

5(a)	B2 for negative (correlation) so as the altitude increases, temperature decreases (B1 for either **negative** (correlation) **or** an explanation that the higher the altitude, the lower the temperature)	For B2 or B1 Accept converse when describing the correlation eg. (negative), as the temperature increases the altitude decreases. Ignore reference to strength of correlation.	**(2)**
(b)	B1 for the line of best fit plotted correctly	Should go through all the points and be drawn with a ruler.	**(1)**
(c)	B1 for −14	B0 if there is no line of best fit. B1ft their line of best fit Allow tolerance on their '−14' of half a small square. For follow through, the gradient of line of best fit must be negative.	**(1)**
(d)	B2 for extrapolating/outside of range for altitude/data only goes up to 10000(m) is not reliable so it is therefore not appropriate (B1 for reference to extrapolation/outside of range for altitude/data only goes up to 10000(m) with no or incorrect conclusion) or B2 for perfect correlation/strong correlation/all points lie on a straight line AND 11000 is close to the range of values for altitude therefore it is appropriate. (B1 for reference to strong correlation/perfect correlation/points lie close to the line of best fit with no or incorrect conclusion)	Allow for beyond/outside of the data plotted or beyond/outside within the table.	**(2)**

WAGOLL

A fair 3-sided spinner numbered 1, 2 and 3 and a fair 4-faced dice numbered 1, 2, 3 and 4 are used in a game.

To play the game, a player spins the spinner once and rolls the dice once. The total score is found by adding the number the spinner lands on and the number the dice lands on.

(a) Complete the sample space diagram to show all the possible total scores.

4-faced dice

3-sided spinner	**1**	**2**	**3**	**4**
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7

(2)

To win the game a player needs to get a total score of at least 6
Chloe plays the game once.

(b) Find the probability that Chloe does **not** win the game.

3/12

(2

(Total for Question 11 is 4 marks

number

11(a) B2 for all 9 correct entries (B1 for at least 5 correct entries) **(2)**

	1	**2**	**3**	**4**
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7

(b) **(2)**

M1 for $\frac{n}{12}$ $1 \leqslant n \leqslant 12$ or for $1 - \frac{n}{12}$ $1 \leqslant n \leqslant 12$

A1 ft for $\frac{9}{12}$ oe

ft their table of values

Allow 75%

If answer seen with no working, award M1A1

Note: If table of values is not fully correct then method and answer may be correct or ft from their table of values.

Foundation Paper 1 Q11

Student A

1 A fair 3-sided spinner numbered 1, 2 and 3 and a fair 4-faced dice numbered 1, 2, 3 and 4 are used in a game.

To play the game, a player spins the spinner once and rolls the dice once. The total score is found by adding the number the spinner lands on and the number the dice lands on.

(a) Complete the sample space diagram to show all the possible total scores.

4-faced dice

3-sided spinner	**1**	**2**	**3**	**4**
1	2	3	4	5
2	3	2 ✗	5	6
3	4	5	6	7

✓

(2)

To win the game a player needs to get a total score of at least 6
Chloe plays the game once.

(b) Find the probability that Chloe does **not** win the game.

✗ $\frac{1}{12}$ ✓

(2)

(Total for Question 11 is 4 marks)

number			
11(a)	B2 for all 9 correct entries (B1 for at least 5 correct entries) **1** **2** **3** **4** **1** 2 3 4 5 **2** 3 4 5 6 **3** 4 5 6 7		**(2)**
(b)	M1 for $\frac{n}{12}$ $1 \leqslant n \leqslant 12$ or for $1-\frac{n}{12}$ $1 \leqslant n \leqslant 12$ A1 ft for $\frac{9}{12}$ oe	ft their table of values Allow 75% If answer seen with no working, award M1A1 Note: If table of values is not fully correct then method and answer may be correct or ft from their table of values.	**(2)**

Foundation Paper 1 Q11

Student B

Foundation Paper 1 Q11

WAGOLL

11 A fair 3-sided spinner numbered 1, 2 and 3 and a fair 4-faced dice numbered 1, 2, 3 and 4 are used in a game.

To play the game, a player spins the spinner once and rolls the dice once. The total score is found by adding the number the spinner lands on and the number the dice lands on.

(a) Complete the sample space diagram to show all the possible total scores.

		4-faced dice			
		1	**2**	**3**	**4**
3-sided spinner	**1**	2	3	4	5
	2	3	4	5	6
	3	4	5	6	7

(2)

To win the game a player needs to get a total score of at least 6 Chloe plays the game once.

$P(6+) = \frac{3}{12}$

(b) Find the probability that Chloe does **not** win the game.

so P (does not win)
$= \frac{9}{12}$

$\frac{9}{12}$

(2)

(Total for Question 11 is 4 marks)

Kyle also collected data about the heights of professional basketball players from 1950 to 1959 and the heights of professional basketball players from 2000 to 2009

The grouped frequency table below gives information about the heights of professional basketball players from 2000 to 2009

Height (h centimetres)	Frequency			
$170 < h \leqslant 180$	12	×	180 =	2160
$180 < h \leqslant 190$	146	×	190 =	27740
$190 < h \leqslant 200$	175	×	200 =	35000
$200 < h \leqslant 210$	323	×	210 =	42000
$210 < h \leqslant 220$	146	×	220 =	32120
$220 < h \leqslant 230$	8	×	230 =	1840
Total	**810**			140860

(Source: *www.kaggle.com*)

The estimate of the mean height for professional basketball players from 1950 to 1959 is calculated to be 190.9 cm to one decimal place.

(e) (i) Calculate an estimate of the mean height of basketball players from 2000 to 2009

140860 / 6

23476.66.. cm

(3)

	• more than half of the basketball players weigh more than the mean		
(e)(i)	M1M1 $\frac{12 \times 175 + 146 \times 185 + 175 \times 195 + 323 \times 205 + 146 \times 215 + 8 \times 225}{810}$ (= 200.79 ...) A1 for 200.79…	M1 for consistent use of fx with x within interval (including end points) and attempt to sum. Use of mid-interval gives $2100 + 27010 + 34125 + 66215 + 31390 + 1800 = 162640$ If multiplications not shown then allow one incorrect product for both M marks. M1 for correct use of fx with x the mid-interval value, attempt to sum and division by 810 A1 for awrt 200.8 OR answer of 201 from correct working	(3)
(e)(ii)	B1ft for e.g. mean height of basketball players has increased (by 9.9 cm)	B1ft for correct comparison of means	(1)

COMMON ITEM
Foundation Paper 1 Q9 Higher Paper 1 Q2 (e only)

Student A

Kyle also collected data about the heights of professional basketball players from 1950 to 1959 and the heights of professional basketball players from 2000 to 2009

The grouped frequency table below gives information about the heights of professional basketball players from 2000 to 2009

Height (h centimetres)	Frequency		
$170 < h \leqslant 180$	12	175	2100
$180 < h \leqslant 190$	146	185	27010
$190 < h \leqslant 200$	175	195	34125
$200 < h \leqslant 210$	323	205	66215
$210 < h \leqslant 220$	146	215	31390
$220 < h \leqslant 230$	8	225	1800
Total	**810**	1200	162640 ✓

(Source: *www.kaggle.com*)

The estimate of the mean height for professional basketball players from 1950 to 1959 is calculated to be 190.9 cm to one decimal place.

(e) (i) Calculate an estimate of the mean height of basketball players from 2000 to 2009

162640 / 1200 ✗

135·53 ✗ cm

(3)

COMMON ITEM
Foundation Paper 1 Q9 Higher Paper 1 Q2 (e only)

Student B

	• more than half of the basketball players weigh more than the mean		
(e)(i)	M1M1 $\frac{12 \times 175 + 146 \times 185 + 175 \times 195 + 323 \times 205 + 146 \times 215 + 8 \times 225}{810}$ (= 200.79 ...) A1 for 200.79…	M1 for consistent use of fx with x within interval (including end points) and attempt to sum. Use of mid-interval gives $2100 + 27010 + 34125 + 66215 + 31390 + 1800 = 162640$ If multiplications not shown then allow one incorrect product for both M marks. M1 for correct use of fx with x the mid-interval value, attempt to sum and division by 810 A1 for awrt 200.8 OR answer of 201 from correct working	(3)
(e)(ii)	B1ft for e.g. mean height of basketball players has increased (by 9.9 cm)	B1ft for correct comparison of means	(1)

Common Item
Foundation P1 Q9 (and Higher P1 Q2)

WAGOLL

Kyle also collected data about the heights of professional basketball players from 1950 to 1959 and the heights of professional basketball players from 2000 to 2009

The grouped frequency table below gives information about the heights of professional basketball players from 2000 to 2009

Height (h centimetres)	Frequency		Midpoint		
$170 < h \leqslant 180$	12	×	175	=	2100
$180 < h \leqslant 190$	146	×	185	=	27010
$190 < h \leqslant 200$	175	×	195	=	34125
$200 < h \leqslant 210$	323	×	205	=	66215
$210 < h \leqslant 220$	146	×	215	=	31390
$220 < h \leqslant 230$	8	×	225	=	1800
Total	**810**		✓		162640

(Source: *www.kaggle.com*)

The estimate of the mean height for professional basketball players from 1950 to 1959 is calculated to be 190.9 cm to one decimal place.

(e) (i) Calculate an estimate of the mean height of basketball players from 2000 to 2009

162640 ÷ 810 = 200.79..

200.8 cm

(3)

(ii) Comment on how the mean height of professional basketball players has changed between the two sets of data.

The mean height of the players has increased by 9.9cm.

(1)

(Total for Question 9 is 10 marks)

Higher Paper 1

Q	Part	Skill tested	Score	Max score	Percent	ALL	1	2	3	4	5	6	7	8	9	ALL	U
Q01	a	Stem and leaf diagram	1.83	3	61	1.83			1	1.42	1.69	1.88	2.01	2.16	2.39	1.83	0.55
Q01	b	Calculate interquartile range (IQR)	0.89	2	45	0.89			0.09	0.3	0.63	0.95	1.18	1.35	1.55	0.89	0.03
Q01	c i	Use median and interquartile range (IQR) to compare data	1.59	3	53	1.59			0.27	0.59	1.09	1.67	2.1	2.45	2.71	1.59	0.17
Q01	c ii	Importance of reliability and validity	0.2	1	20	0.2			0.04	0.09	0.14	0.2	0.24	0.31	0.46	0.2	0.01
Q01	d	Techniques used to deal with problems with collected data	0.25	1	25	0.25			0.07	0.09	0.19	0.23	0.3	0.4	0.58	0.25	0.02
Q02	a	Terms used to describe different types of data	0.69	1	69	0.69			0.33	0.45	0.58	0.72	0.82	0.9	0.97	0.69	0.25
Q02	b	Histograms (equal class width) - completing a table	1.76	2	88	1.76			1.44	1.64	1.74	1.8	1.83	1.87	1.94	1.76	0.75
Q02	c	Histograms (equal class width) - completing a histogram	1.52	2	76	1.52			1.2	1.42	1.52	1.55	1.57	1.6	1.67	1.52	0.57
Q02	d	Interpret a distribution of data in term of skewness	0.1	1	10	0.1				0.01	0.01	0.04	0.1	0.25	0.51	0.1	
Q02	e i	Calculate an estimate of the mean	2.07	3	69	2.07			0.48	0.98	1.69	2.29	2.66	2.83	2.9	2.07	0.2
Q02	e ii	Compare data sets using appropriate measure of central tendency	0.84	1	84	0.84			0.37	0.61	0.78	0.9	0.96	0.98	0.99	0.84	0.22
Q03	a	Hypothesis testing	0.51	1	51	0.51			0.11	0.22	0.38	0.53	0.64	0.75	0.86	0.51	0.03
Q03	b	Interpret seasonal and cyclic trends in context	1.39	2	70	1.39			0.47	0.9	1.25	1.48	1.64	1.72	1.86	1.39	0.23
Q03	c	Comment of appropriateness of using moving averages	0.59	1	59	0.59			0.15	0.33	0.44	0.61	0.74	0.85	0.92	0.59	0.06
Q03	d	Compare seasonal and cyclic trends in context	0.43	1	43	0.43			0.06	0.2	0.33	0.43	0.53	0.63	0.77	0.43	0.01
Q03	e	Time series - Plotting	2.01	3	67	2.01			0.63	1.27	1.81	2.15	2.37	2.54	2.73	2.01	0.22
Q03	f	Interpret seasonal and cyclic trends in context	1.37	2	69	1.37			0.36	0.71	1.13	1.47	1.73	1.85	1.94	1.37	0.15
Q04		Strategies to mitigate issues during the statistical enquiry cycle	2.04	5	41	2.04			0.41	0.8	1.42	2	2.59	3.13	3.91	2.04	0.14
Q05	a	Appropriateness of SRCC	0.09	1	9	0.09			0.01	0.02	0.05	0.06	0.11	0.15	0.27	0.09	
Q05	b	Apply formula to determine Spearman's rank correlation coefficient	2.88	5	58	2.88			0.34	0.9	1.94	3.15	4	4.39	4.63	2.88	0.13
Q06	a	Use chain base index numbers in context	1.22	2	61	1.22			0.1	0.35	0.8	1.35	1.72	1.88	1.97	1.22	0.03
Q06	b i	Calculate geometric mean	0.95	2	48	0.95			0.03	0.09	0.4	0.92	1.45	1.81	1.95	0.95	
Q06	b ii	Interpret geometric mean	0.7	2	35	0.7			0.02	0.07	0.29	0.65	1.04	1.35	1.63	0.7	
Q07	a	Capture recapture formula to estimate the size of a population	1.28	2	64	1.28			0.14	0.44	0.98	1.46	1.68	1.82	1.94	1.28	0.03
Q07	b	Appropriateness of capture recapture	1.77	3	59	1.77			0.64	1.09	1.55	1.86	2.08	2.25	2.53	1.77	0.18
Q08	a	Tree diagrams	0.86	1	86	0.86			0.51	0.7	0.84	0.9	0.95	0.97	0.98	0.86	0.23
Q08	b	Calculate conditional probability	0.16	2	8	0.16					0.01	0.03	0.11	0.37	1.08	0.16	
Q08	c i	Calculate relative risk	0.27	2	14	0.27			0.01	0.03	0.04	0.12	0.31	0.64	1.16	0.27	
Q08	c ii	Interpret relative risk	0.21	1	21	0.21				0.02	0.04	0.1	0.24	0.49	0.8	0.21	
Q08	d	Interpret relative risk	0.13	2	7	0.13				0.01	0.02	0.06	0.13	0.25	0.63	0.13	
Q09	a	Determine skewness	1.69	2	85	1.69			0.79	1.26	1.64	1.81	1.9	1.95	1.96	1.69	0.38
Q09	b	Use mean and standard deviation to compare data	2.27	5	45	2.27			0.3	0.82	1.51	2.23	2.92	3.68	4.35	2.27	0.09
Q09	c	Justify the rationale for selecting appropriate types of average	0.27	1	27	0.27			0.02	0.05	0.12	0.21	0.36	0.53	0.72	0.27	
Q10	a	Determine line of best fit	0.14	1	14	0.14					0.01	0.03	0.14	0.38	0.75	0.14	
Q10	b	Determine line of best fit	0.26	3	9	0.26			0.01	0.03	0.08	0.12	0.24	0.52	1.21	0.26	
Q10	c i	Characteristics of a normal distribution	0.35	1	35	0.35			0.01	0.03	0.12	0.27	0.51	0.72	0.89	0.35	
Q10	c ii	Characteristics of a normal distribution	0.27	2	14	0.27				0.01	0.03	0.11	0.34	0.72	1.06	0.27	
Q11	a	Determine factors that may lead to bias	0.54	1	54	0.54			0.13	0.23	0.37	0.52	0.69	0.84	0.96	0.54	0.05
Q11	b	Determine factors that may lead to bias	0.21	2	11	0.21			0.04	0.05	0.09	0.15	0.26	0.4	0.7	0.21	
Q11	c	Use stratification and know when this is appropriate before sampling	0.25	2	13	0.25				0.01	0.03	0.12	0.29	0.59	1	0.25	
			36.85	80		36.85	0	0	10.58	18.24	27.78	37.13	45.48	53.27	62.83	36.85	4.73

The scores for Vitality Roses netball team for the 2017 season are listed below.

55	61	72	62	52	74	56	49	75	66
64	85	55	52	55	94	61	46	74	54

(Source: *www.englandnetball.co.uk*)

(a) Use the data to complete the back-to-back stem and leaf diagram.

2017 season		2018 season
	1	9
	2	5 6 9
9 6	3	5 7 9
22 4 55 5 6	4	3 3 5 5 5 7
1 1 2 4 6	5	0 2 2 4 6 8 8
2 4 4 5	6	4 5 6
5	7	0 2
4	8	
	9	0

Key: 4|9 = 49

(3)

(b) Work out the interquartile range of the scores for the 2018 season.

58 - 39

17

(2)

number			
1(a)	B2 for correctly completing the 2017 season on the stem and leaf diagram B1 for a suitable key Key: 4 | 9 | 0 represents a score of 94 in the 2017 season and 90 in the 2018 season	2017 season | 2018 season 1 | 9 2 | 5 6 9 3 | 5 7 9 9 6 | 4 | 3 3 5 5 5 7 6 5 5 5 4 2 2 | 5 | 0 2 2 4 6 8 8 6 4 2 1 1 | 6 | 4 5 6 5 4 4 2 | 7 | 0 2 5 | 8 | 4 | 9 | 0	B2 for a fully correct back-to-back stem and leaf diagram (B1 for ordered diagram with at most 2 errors or for unordered diagram) 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 2017 and 2018 or it must be clear how this is interpreted for the two sides. e.g. 4 | 9 represents a score of 94 in the 2017 (season) and 9|0 represents a score of 90 in the 2018 (season) or |1|9 represents 19, 4|9| represents 94.	(3)
(b)	M1 for $58 - 39$ A1 19	M1 for $58 - k$ or $k - 39$ or for both 39 and 58 identified or for $59.5 - k$ or $k - 38.5$ or for both 38.5 and 59.5 identified (leads to IQR=21) or for $58 - k$ or $k - 38$ or for both 38 and 58 identified (leads to IQR=20) A1 accept 20 or 21 If working is seen then it must be correct and consistent with their answer for award of M1A1. Accept if answer given in table.	(2)

Higher Paper 1 Q1 (a / b)

Student A

The scores for Vitality Roses netball team for the 2017 season are listed below.

55	61	72	62	52	74	56	49	75	66
64	85	55	52	55	94	61	46	74	54

(Source: *www.englandnetball.co.uk*)

(a) Use the data to complete the back-to-back stem and leaf diagram.

2017 season		2018 season
	1	9
	2	5 6 9
	3	5 7 9
96	4	3 3 5 5 5 7
6 5 5 5 4 2 2	5	0 2 2 4 6 8 8
6 4 2 1 1	6	4 5 6
5 4 4 2	7	0 2
5	8	
4	9	0

Key: 2 | 5 | 1 = 52 ✓ 1 | 4 | 9 = 49

(3)

(b) Work out the interquartile range of the scores for the 2018 season.

✓✓

19

(2)

Number			
1(a)	B2 for correctly completing the 2017 season on the stem and leaf diagram B1 for a suitable key Key: 4 | 9 | 0 represents a score of 94 in the 2017 season and 90 in the 2018 season		B2 for a fully correct back-to-back stem and leaf diagram (B1 for ordered diagram with at most 2 errors or for unordered diagram) 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 2017 and 2018 or it must be clear how this is interpreted for the two sides. e.g. 4 | 9 represents a score of 94 in the 2017 (season) and 9|0 represents a score of 90 in the 2018 (season) or |1|9 represents 19, 4|9| represents 94.	(3)
(b)	M1 for $58 - 39$ A1 19		M1 for $58 - k$ or $k - 39$ or for both 39 and 58 identified or for $59.5 - k$ or $k - 38.5$ or for both 38.5 and 59.5 identified (leads to IQR=21) or for $58 - k$ or $k - 38$ or for both 38 and 58 identified (leads to IQR=20) A1 accept 20 or 21 If working is seen then it must be correct and consistent with their answer for award of M1A1. Accept if answer given in table.	(2)

2017 season		2018 season
	1	9
	2	5 6 9
	3	5 7 9
9 6	4	3 3 5 5 5 7
6 5 5 5 4 2 2	5	0 2 2 4 6 8 8
6 4 2 1 1	6	4 5 6
5 4 4 2	7	0 2
5	8	
4	9	0

Higher Paper 1 Q1 (a / b)

Student B

The table gives information about the scores for the 2017 and 2018 seasons.

	Median	Interquartile range
2017 season	61	18.5
2018 season	48.5	

Naomi thinks that the results show that Vitality Roses had a higher and more consistent set of scores in the 2018 season than in the 2017 season.

(c) (i) Do the statistics support this conclusion?
You must give reasons for your answer.

Yes the median in 2017 was higher so is more consistent.

(3)

(ii) Give a limitation of using the data provided to comment on whether or not Vitality Roses has improved as a team between the 2017 and 2018 seasons.

The IQR is missing

(1)

Naomi plans to process the data further.

She suggests checking for outliers, by performing calculations, before doing any more processing.

(d) Comment on the appropriateness of this suggestion.

yes outliers can affect the data

(1)

(Total for Question 1 is 10 marks)

(c)(i)	B1 for the median in the 2017 season is higher than the median in the 2018 season B1ft for the interquartile range for the 2017 season is lower than the interquartile range in the 2018 season B1ft (dep) so the team had higher scores and were more consistent in the 2017 season (so no Naomi is not correct)	B1 for correct comparison of medians. Accept use of average rather than median. Does not need to be in context. Allow e.g. median for 2017 is higher. B1ft for correct comparison of interquartile range. Must be correct for their IQR. Does not need to be in context. Allow e.g. IQR for 2017 is lower. B1ft for assessing the appropriateness of the conclusion. Dep on B1B1 scored (question indicates reasons required) Final B mark needs to have contextual interpretation seen in the answer e.g. …so higher scores, … so more consistent. Do not award the 3rd B mark if the reasoning is linked with the wrong statistic e.g. linking median with consistency.	(3)
(ii)	B1 for giving a limitation of using the data provided to comment on team improvement e.g. • we don't know if the matches were played against the same teams in both seasons • we don't know what the scores were for the opposition teams • opposition team performance may also have changed	B1 for a limitation of using the data provided for the conclusion Do not accept: • team members might have changed Accept: • we only know the scores, we don't know the results of the matches. • we don't know how many times they have won/lost each season.	(1)

(d)	B1 for e.g. • appropriate as if there are outliers in the data then it might impact further diagrams or calculations • appropriate as there appear to be outliers in the data / 90 is much higher than the other values for 2018 season • not appropriate as there is nothing to suggest that the data is not genuine • not appropriate as using all data would be fully representative	B1 for assessing the appropriateness of checking for outliers by calculation Accept 'yes' for appropriate and 'no' for not appropriate. Allow reference to specific diagrams or calculations, but not median or interquartile range as these have already been calculated. Do not accept as reasons: • …making it more accurate/precise • …make results fair • …an outlier can affect the data with no other comment / clarification • …using median/IQR which has already removed outliers (this has already been done) • …they should use all the data (needs further explanation)	(1)

Higher Paper 1 Q1 (c / d)

Student A

The table gives information about the scores for the 2017 and 2018 seasons.

	Median	Interquartile range
2017 season	61	18.5
2018 season	48.5	19

Naomi thinks that the results show that Vitality Roses had a higher and more consistent set of scores in the 2018 season than in the 2017 season.

(c) (i) Do the statistics support this conclusion?
You must give reasons for your answer.

No, the median is lower in 2018 and the IQR is higher.

(3)

(ii) Give a limitation of using the data provided to comment on whether or not Vitality Roses has improved as a team between the 2017 and 2018 seasons.

it depends. We don't know if they won the games, just the scores.

(1)

Naomi plans to process the data further.

She suggests checking for outliers, by performing calculations, before doing any more processing.

(d) Comment on the appropriateness of this suggestion.

No, they should use all the data

(1)

(Total for Question 1 is 10 marks)

(c)(i)	B1 for the median in the 2017 season is higher than the median in the 2018 season B1ft for the interquartile range for the 2017 season is lower than the interquartile range in the 2018 season B1ft (dep) so the team had higher scores and were more consistent in the 2017 season (so no Naomi is not correct)	B1 for correct comparison of medians. Accept use of average rather than median. Does not need to be in context. Allow e.g. median for 2017 is higher. B1ft for correct comparison of interquartile range. Must be correct for their IQR. Does not need to be in context. Allow e.g. IQR for 2017 is lower. B1ft for assessing the appropriateness of the conclusion. Dep on B1B1 scored (question indicates reasons required) Final B mark needs to have contextual interpretation seen in the answer e.g. …so higher scores, … so more consistent. Do not award the 3rd B mark if the reasoning is linked with the wrong statistic e.g. linking median with consistency.	(3)
(ii)	B1 for giving a limitation of using the data provided to comment on team improvement e.g. • we don't know if the matches were played against the same teams in both seasons • we don't know what the scores were for the opposition teams • opposition team performance may also have changed	B1 for a limitation of using the data provided for the conclusion Do not accept: • team members might have changed Accept: • we only know the scores, we don't know the results of the matches. • we don't know how many times they have won/lost each season.	(1)

(d)	B1 for e.g. • appropriate as if there are outliers in the data then it might impact further diagrams or calculations • appropriate as there appear to be outliers in the data / 90 is much higher than the other values for 2018 season • not appropriate as there is nothing to suggest that the data is not genuine • not appropriate as using all data would be fully representative	B1 for assessing the appropriateness of checking for outliers by calculation Accept 'yes' for appropriate and 'no' for not appropriate. Allow reference to specific diagrams or calculations, but not median or interquartile range as these have already been calculated. Do not accept as reasons: • …making it more accurate/precise • …make results fair • …an outlier can affect the data with no other comment / clarification • …using median/IQR which has already removed outliers (this has already been done) • …they should use all the data (needs further explanation)	(1)

Higher Paper 1 Q1 (c / d)

Student B

Higher Paper 1 Q1

WAGOLL

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 The scores for Vitality Roses netball team for the 2017 season are listed below.

55	61	72	62	52	74	56	49	75	66
64	85	55	52	55	94	61	46	74	54

(Source: *www.englandnetball.co.uk*)

(a) Use the data to complete the back-to-back stem and leaf diagram.

2017 season		2018 season
	1	4
	2	4 4 5
	3	4 4 9
9 6	4	0 0 1 2 3 4
6 5 5 5 4 2 2	5	0 2 3 3 4 6 8 8
6 4 2 1 1	6	0 3 6
5 4 4 2	7	0 2
5	8	
4	9	0

Key: 4 \| 9 \| 0

49 in 2017 90 in 2018

(3)

(b) Work out the interquartile range of the scores for the 2018 season.

58 - 39 = 19

19

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

The table gives information about the scores for the 2017 and 2018 seasons.

	Median	Interquartile range
2017 season	61 ↑	18.5
2018 season	48.5	19

Naomi thinks that the results show that Vitality Roses had a higher and more consistent set of scores in the 2018 season than in the 2017 season.

(c) (i) Do the statistics support this conclusion?
You must give reasons for your answer.

No, the statistics do not support the conclusion. The median in 2017 was higher than in 2018 which suggests the scores were on average higher in 2017 and the IQR in 2017 was lower than 2018 which means they were more consistent in 2017

(3)

(ii) Give a limitation of using the data provided to comment on whether or not Vitality Roses has improved as a team between the 2017 and 2018 seasons.

We don't know anything about the teams these matches were played against.. the oppositions performance may also have changed

(1)

Naomi plans to process the data further.

She suggests checking for outliers, by performing calculations, before doing any more processing.

(d) Comment on the appropriateness of this suggestion.

It would be appropriate to do this as there appears to be an outlier in 2018.

(1)

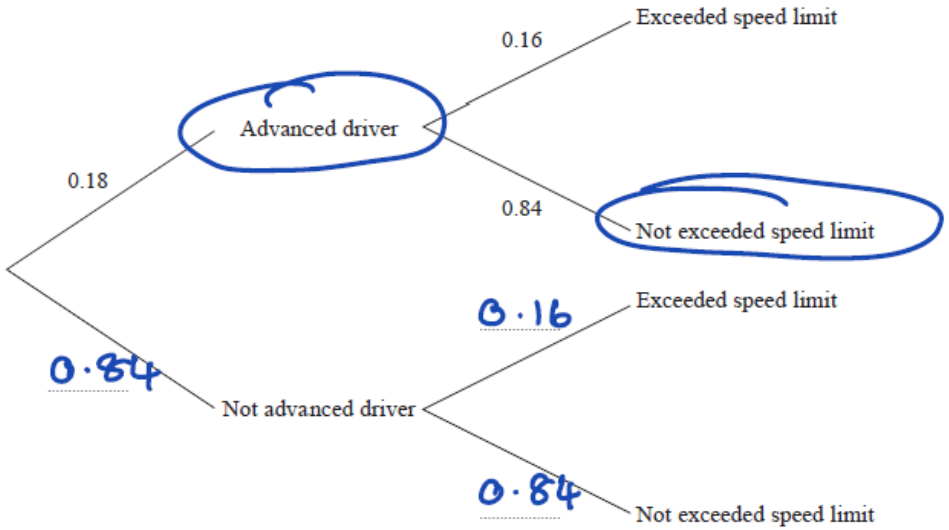
(Total for Question 1 is 10 marks)

8 Ian collects information about whether drivers have had advanced driver training and whether they have exceeded the speed limit in the last month.

He finds that 18% of the drivers he surveyed have had advanced driver training and that 16% of these drivers have exceeded the speed limit in the last month.

For the drivers who have not had advanced driver training 26% of them have exceeded the speed limit in the last month.

(a) Complete the probability tree diagram for this information.



(1)

(b) Calculate P(Advanced driver | Not exceeded speed limit).

(

0·18 + 0·84 = 1·02

.....

(2)

number			
8(a)	B1 for 0.82, 0.26, 0.74 oe in correct places on tree diagram	B1 for correct probabilities on the tree diagram	(1)
(b)	M1 for $\frac{0.18\times0.84}{0.18\times0.84+"0.82"\times"0.74"}$ A1 for 0.199(4…)	M1 for correct method to calculate conditional probability. Allow use of their figures from (a) as long as these are probabilities (between 0 and 1). A1 for awrt 0.20 Accept 0.2 with working.	(2)

Higher Paper 1 Q8 (a / b)

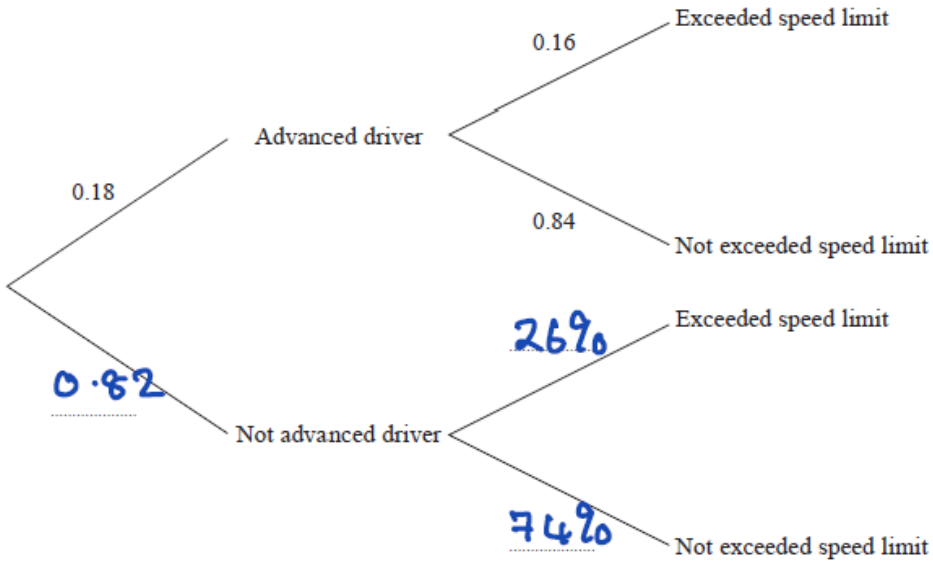
Student A

8 Ian collects information about whether drivers have had advanced driver training and whether they have exceeded the speed limit in the last month.

He finds that 18% of the drivers he surveyed have had advanced driver training and that 16% of these drivers have exceeded the speed limit in the last month.

For the drivers who have not had advanced driver training 26% of them have exceeded the speed limit in the last month.

(a) Complete the probability tree diagram for this information.



(1)

(b) Calculate P(Advanced driver | Not exceeded speed limit).

.....

(2)

number			
8(a)	B1 for 0.82, 0.26, 0.74 oe in correct places on tree diagram	B1 for correct probabilities on the tree diagram	(1)
(b)	M1 for $\frac{0.18\times0.84}{0.18\times0.84+"0.82"\times"0.74"}$ A1 for 0.199(4…)	M1 for correct method to calculate conditional probability. Allow use of their figures from (a) as long as these are probabilities (between 0 and 1). A1 for awrt 0.20 Accept 0.2 with working.	(2)

Higher Paper 1 Q8 (a / b)

Student B

(c) (i) Calculate the relative risk of exceeding the speed limit for the not advanced drivers compared to the advanced drivers.

0·26 / 0·16

26% - 16%

10%

(2)

(ii) Interpret this relative risk in context.

10% more likely to speed.

(1)

Ian researches the relative risk of exceeding the speed limit for newly qualified drivers compared to drivers who are not newly qualified.

He finds that in 2019 the relative risk was 0.95 and in 2020 the relative risk was 0.87

Ian concludes that the proportion of newly qualified drivers exceeding the speed limit has decreased between 2019 and 2020

(d) Do the statistics support Ian's conclusion?
You must give a reason for your answer.

its not possible to say, it could be this or it could be that the others' propotion has increased.

(2)

(Total for Question 8 is 8 marks)

		Accept 0.2 with working.	
(c)(i)	M1 for $\frac{"0.26"}{0.16}$ A1 for 1.625	M1 for correct method to calculate relative risk A1 cao (allow 1.63)	(2)
(c)(ii)	B1ft for the risk of a driver who has not had advanced training breaking the speed limit is "1.625" times/"162.5%" of the risk of a driver who has had advanced training breaking the speed limit oe	B1ft for correct interpretation of their relative risk in context Allow use of their figure rounded in comment e.g. 1.63, 1.6 Allow e.g. the risk of a driver who has not had advanced training breaking the speed limit is "**62.5%" more than** the risk of a driver who has had advanced training breaking the speed limit Do not allow for incorrect interpretation e.g. the risk of a driver who has not had advanced training breaking the speed limit is **"162.5%" more than** risk of a driver who has had advanced training breaking the speed limit – this is not equivalent to 162.5% of the risk of…. Condone if a correct statement seen followed by an incorrect attempt to write in a different form.	(1)

(d)	B2 for e.g. no / not possible to tell from the data, it could be that the proportion of newly qualified drivers speeding has reduced or it could be that the proportion of not newly qualified drivers speeding has increased (B1 for e.g. it could be that the proportion of newly qualified drivers speeding has reduced or it could be that the proportion of not newly qualified drivers speeding has increased)	B2 for a complete reason assessing the appropriateness of the conclusion with statistical reason Accept for B2 no as the relative risk may have decreased due to drivers who are not newly qualified exceeding the speed limit more than before. (B1 for correct interpretation of the change in relative risk with incorrect or no conclusion) Accept for B1 the relative risk may have decreased due to drivers who are not newly qualified exceeding the speed limit more than before with no or incorrect conclusion.	(2)

Higher Paper 1 Q8 (c / d)

Student A

(c) (i) Calculate the relative risk of exceeding the speed limit for the not advanced drivers compared to the advanced drivers.

0·18 / 0·16

1·125

(2)

(ii) Interpret this relative risk in context.

The risk of a driver who has not had advanced training breaking the speed limit is 1·125 times the risk of a driver who had had the training

(1)

Ian researches the relative risk of exceeding the speed limit for newly qualified drivers compared to drivers who are not newly qualified.

He finds that in 2019 the relative risk was 0.95 and in 2020 the relative risk was 0.87

Ian concludes that the proportion of newly qualified drivers exceeding the speed limit has decreased between 2019 and 2020

(d) Do the statistics support Ian's conclusion?
You must give a reason for your answer.

Yes 0·95 > 0·87

(2)

(Total for Question 8 is 8 marks)

		Accept 0.2 with working.	
(c)(i)	M1 for $\frac{"0.26"}{0.16}$ A1 for 1.625	M1 for correct method to calculate relative risk A1 cao (allow 1.63)	(2)
(c)(ii)	B1ft for the risk of a driver who has not had advanced training breaking the speed limit is "1.625" times/"162.5%" of the risk of a driver who has had advanced training breaking the speed limit oe	B1ft for correct interpretation of their relative risk in context Allow use of their figure rounded in comment e.g. 1.63, 1.6 Allow e.g. the risk of a driver who has not had advanced training breaking the speed limit is "**62.5%" more than** the risk of a driver who has had advanced training breaking the speed limit Do not allow for incorrect interpretation e.g. the risk of a driver who has not had advanced training breaking the speed limit is **"162.5%" more than** risk of a driver who has had advanced training breaking the speed limit – this is not equivalent to 162.5% of the risk of…. Condone if a correct statement seen followed by an incorrect attempt to write in a different form.	(1)

(d)	B2 for e.g. no / not possible to tell from the data, it could be that the proportion of newly qualified drivers speeding has reduced or it could be that the proportion of not newly qualified drivers speeding has increased (B1 for e.g. it could be that the proportion of newly qualified drivers speeding has reduced or it could be that the proportion of not newly qualified drivers speeding has increased)	B2 for a complete reason assessing the appropriateness of the conclusion with statistical reason Accept for B2 no as the relative risk may have decreased due to drivers who are not newly qualified exceeding the speed limit more than before. (B1 for correct interpretation of the change in relative risk with incorrect or no conclusion) Accept for B1 the relative risk may have decreased due to drivers who are not newly qualified exceeding the speed limit more than before with no or incorrect conclusion.	(2)

Higher Paper 1 Q8 (c / d)

Student B

Higher Paper 1 Q8

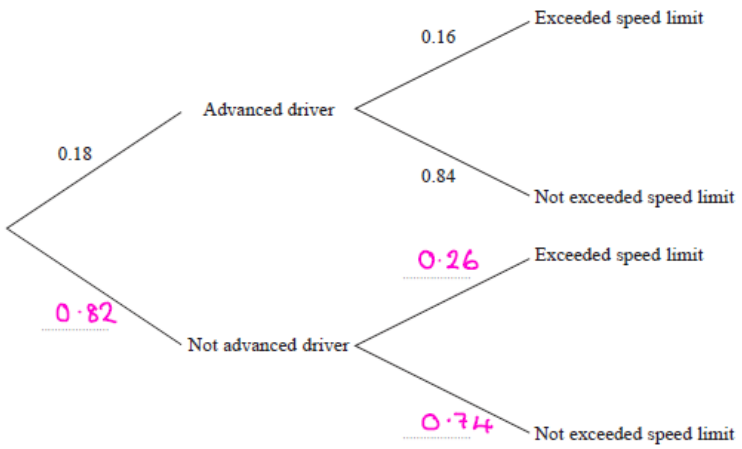
WAGOLL

8 Ian collects information about whether drivers have had advanced driver training and whether they have exceeded the speed limit in the last month.

He finds that 18% of the drivers he surveyed have had advanced driver training and that 16% of these drivers have exceeded the speed limit in the last month.

For the drivers who have not had advanced driver training 26% of them have exceeded the speed limit in the last month.

(a) Complete the probability tree diagram for this information.



(1)

(b) Calculate P(Advanced driver | Not exceeded speed limit).

$$\frac{0.18 \times 0.84}{0.18 \times 0.84 + 0.82 \times 0.74}$$

$$= \frac{0.1512}{0.1512 + 0.6068}$$

$$= \frac{0.1512}{0.758} = 0.19947\ldots$$

0.2 (1dp)

(2)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

(c) (i) Calculate the relative risk of exceeding the speed limit for the not advanced drivers compared to the advanced drivers.

$$\frac{0.26}{0.16} = 1.625$$

1.625

(2)

(ii) Interpret this relative risk in context.

The risk of a driver who has not had advanced training breaking the speed limit is 1.625 times the risk of a driver who has had the training breaking the speed limit

(1)

Ian researches the relative risk of exceeding the speed limit for newly qualified drivers compared to drivers who are not newly qualified.

He finds that in 2019 the relative risk was 0.95 and in 2020 the relative risk was 0.87

Ian concludes that the proportion of newly qualified drivers exceeding the speed limit has decreased between 2019 and 2020

(d) Do the statistics support Ian's conclusion?
You must give a reason for your answer.

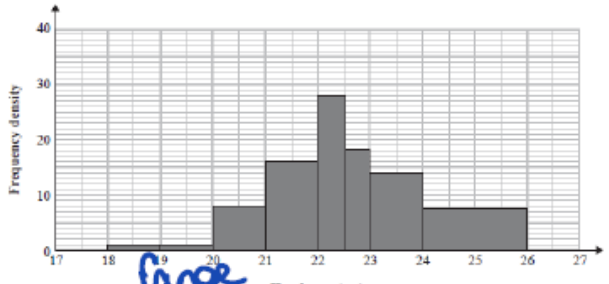
It is not possible to say, so no Ians conclusion is not supported. It could be that the propotion of newly qualified drivers has reduced or not newly qualified drivers speeding has increased.

(2)

(Total for Question 8 is 8 marks)

9 Ligita collected data on the handspans, in cm, of men and the handspans, in cm, of women.
She processed her data using statistical software, here are her results.

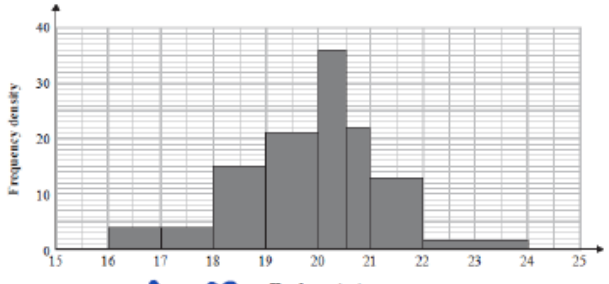
Male handspans



range 7.5

Minimum	18.0	Mean	22.3	Standard deviation	1.44
Maximum	25.5	Median	22.0	Skewness	0.630
Number of male handspans	78				

Female handspans



range 4

Minimum	16.0	Mean	19.6	Standard deviation	1.33
Maximum	23.0	Median	20.0		
Number of female handspans	89				

(Source: nrich.maths.org/)

skewness = 0.902

(a) Calculate the skew of the female handspan data.

$$\frac{3 \times (19.6 - 20)}{1.33}$$

0.902

(2)

number			
9(a)	M1 for $\frac{3\times(19.6-20.0)}{1.33}$ A1 for -0.902	M1 for a method to calculate skewness A1 for awrt -0.90 Accept -0.9 or $-\frac{120}{133}$	(2)

Higher Paper 1 Q9 (a) Student A

(2)

(b) Compare in context the distribution of male handspans and the distribution of female handspans.

Male handspans are bigger, so men have bigger hands.

(5)

Ligita plans to use the mean for the males and the mean for the females to find all of the handspans.

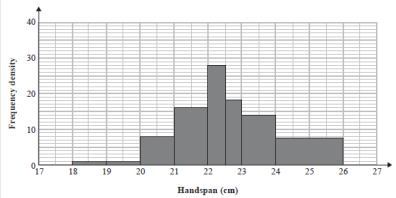
(c) Explain why she should use a weighted mean for this calculation.

(1)

(Total for Question 9 is 8 marks)

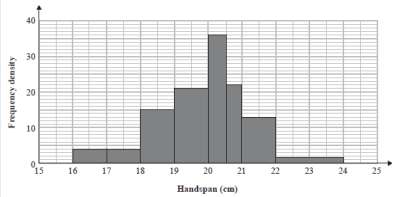
9 Ligita collected data on the handspans, in cm, of men and the handspans, in cm, of women.
She processed her data using statistical software, here are her results.

Male handspans



Minimum	18.0	Mean	22.3	Standard deviation	1.44
Maximum	25.5	Median	22.0	Skewness	0.630
Number of male handspans	78				

Female handspans



Minimum	16.0	Mean	19.6	Standard deviation	1.33
Maximum	23.0	Median	20.0		
Number of female handspans	89				

(Source: mathcct.colano.edu/)

(b)	B1B1B1B1B1 for each of five comparisons or contextual interpretations **Comparison of average** e.g. • mean for male handspans is greater • median for male handspans is greater **Interpretation** e.g. • …on average males have greater / wider handspans **Comparison of measure of spread** e.g. • standard deviation for male handspans is greater • range for male handspans is greater **Interpretation** e.g. • …male handspans are more varied / less consistent / more spread out **Comparison of skew** e.g. • male handspan data is positively skewed, female handspan data is negatively skewed **Interpretation** e.g. • …neither male nor female handspans are normally distributed • …for male handspans the mean handspan is greater than the median handspan for female handspans the mean handspan is less than the median handspan • …for male handspans the spread of handspans greater than the median handspan is greater than the spread of handspans less than the median, but for female handspans the spread of handspans greater	B1B1B1B1B1 for each of five comparisons or contextual interpretations B1 for comparison of measure of central tendency. Award mark for comparison of means **or** for comparison of medians. B1(dep) for correct interpretation of comparison of central tendency Accept e.g. males have bigger hands. B1 for comparison of a measure of spread Condone wider range. B1(dep) for correct interpretation of comparison of a measure of spread Do not allow more distributed or wider for contextual interpretation of spread. B1ft for comparison of skew Follow through their (a). Allow correct or ft (may judge from (a) or histogram). B1(dep) for correct interpretation of comparison of skew Follow through their (a) For comparisons correct statistical terminology is required i.e. mean, median, standard deviation, range, skew.	(5)

	than the median handspan is less than the spread of handspans less than the median • … more than half of the males have handspans lower than the mean for male handspans, however more than half of the females have handspans greater than the mean for female handspans.	Condone use of 'average' when comparing mean or median if values are given.	
(c)	B1 for e.g. • reference to unequal numbers of male and female handspan measurements • sample size for females is greater • sample sizes are different	B1 for correct justification of the appropriateness of using weighted mean	(1)

Higher Paper 1 Q9 (b/c)

Student A

(2)

(b) Compare in context the distribution of male handspans and the distribution of female handspans.

Means for males is greater.

Median for males is greater.

so on average the mens handspans are bigger.

(5)

Ligita plans to use the mean for the males and the mean for the females to find all of the handspans.

(c) Explain why she should use a weighted mean for this calculation.

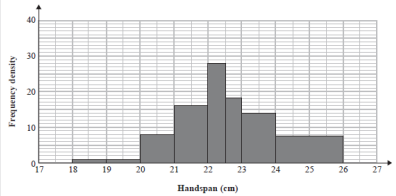
Different numbers are included ie. there were 89 females and 79 males

(1)

(Total for Question 9 is 8 marks)

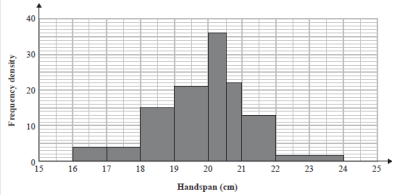
9 Ligita collected data on the handspans, in cm, of men and the handspans, in cm, of women.
She processed her data using statistical software, here are her results.

Male handspans



Minimum	18.0	Mean	22.3	Standard deviation	1.44
Maximum	25.5	Median	22.0	Skewness	0.630
Number of male handspans	78				

Female handspans



Minimum	16.0	Mean	19.6	Standard deviation	1.33
Maximum	23.0	Median	20.0		
Number of female handspans	89				

(Source: *mathcct.colano.edu*)

(b)	B1B1B1B1B1 for each of five comparisons or contextual interpretations **Comparison of average** e.g. • mean for male handspans is greater • median for male handspans is greater **Interpretation** e.g. • …on average males have greater / wider handspans **Comparison of measure of spread** e.g. • standard deviation for male handspans is greater • range for male handspans is greater **Interpretation** e.g. • …male handspans are more varied / less consistent / more spread out **Comparison of skew** e.g. • male handspan data is positively skewed, female handspan data is negatively skewed **Interpretation** e.g. • …neither male nor female handspans are normally distributed • …for male handspans the mean handspan is greater than the median handspan for female handspans the mean handspan is less than the median handspan • …for male handspans the spread of handspans greater than the median handspan is greater than the spread of handspans less than the median, but for female handspans the spread of handspans greater	B1B1B1B1B1 for each of five comparisons or contextual interpretations B1 for comparison of measure of central tendency. Award mark for comparison of means **or** for comparison of medians. B1(dep) for correct interpretation of comparison of central tendency Accept e.g. males have bigger hands. B1 for comparison of a measure of spread Condone wider range. B1(dep) for correct interpretation of comparison of a measure of spread Do not allow more distributed or wider for contextual interpretation of spread. B1ft for comparison of skew Follow through their (a). Allow correct or ft (may judge from (a) or histogram). B1(dep) for correct interpretation of comparison of skew Follow through their (a) For comparisons correct statistical terminology is required i.e. mean, median, standard deviation, range, skew.	(5)

	than the median handspan is less than the spread of handspans less than the median • … more than half of the males have handspans lower than the mean for male handspans, however more than half of the females have handspans greater than the mean for female handspans.	Condone use of 'average' when comparing mean or median if values are given.	
(c)	B1 for e.g. • reference to unequal numbers of male and female handspan measurements • sample size for females is greater • sample sizes are different	B1 for correct justification of the appropriateness of using weighted mean	(1)

Higher Paper 1 Q9 (b/c)

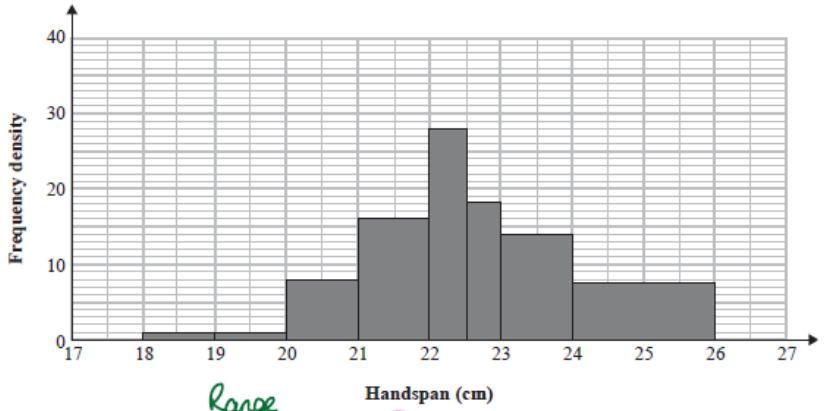
Student B

Higher Paper 1 Q9

WAGOLL

9 Ligita collected data on the handspans, in cm, of men and the handspans, in cm, of women.
She processed her data using statistical software, here are her results.

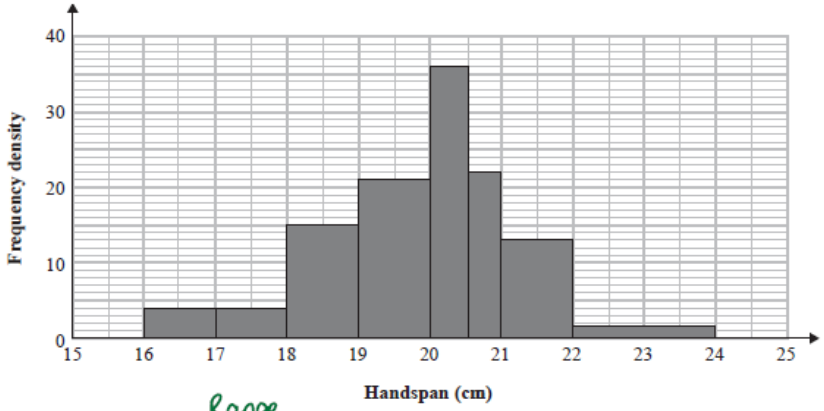
Male handspans



Range 7·5

Minimum	18.0	Mean	22.3	Standard deviation	1.44
Maximum	25.5	Median	22.0	Skewness	0.630
Number of male handspans	78				+ve

Female handspans



Range 7

Minimum	16.0	Mean	19.6	Standard deviation	1.33
Maximum	23.0	Median	20.0	skewness	-0·902
Number of female handspans	89				-ve

(Source: *mathsci.solano.edu/*)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

(a) Calculate the skew of the female handspan data.

$$\frac{3 \times (19.6 - 20)}{1.33}$$

$$= -0.902$$

-0·902

(2)

(b) Compare in context the distribution of male handspans and the distribution of female handspans.

The male mean handspan is greater which means on average the males have wider handspans

The females range is smaller so the handspan for females is more consistent

Male handspans are positively skewed and females is negatively skewed. So for the males the mean handspan is greater than the median and for females the mean handspan is less than the median.

(5)

Ligita plans to use the mean for the males and the mean for the females to find all of the handspans.

(c) Explain why she should use a weighted mean for this calculation.

The sample sizes are different.

(1)

(Total for Question 9 is 8 marks)

Paper 2

Pearson
Edexcel

Foundation Paper 2

FP2

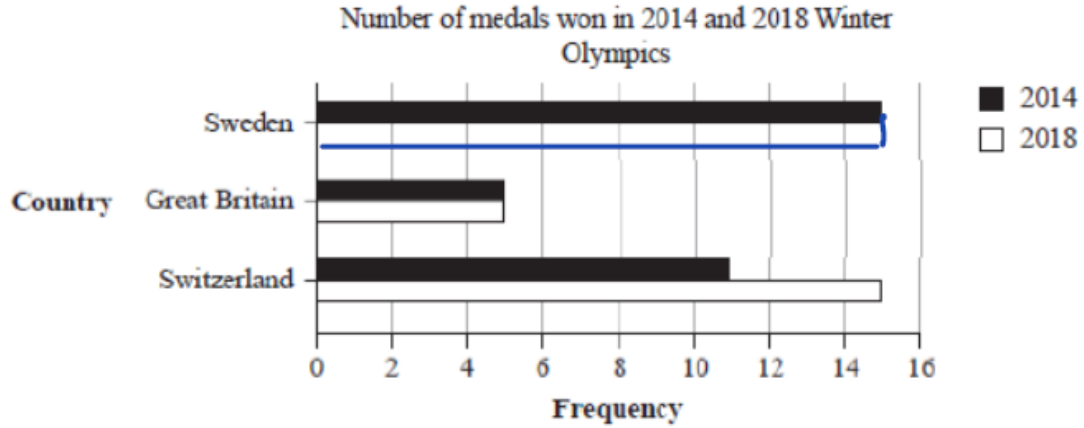
Edexcel averages:

Q	Part	Skill tested	Score	Max score	Percent	ALL	1	2	3	4	5	6	7	8	9	ALL	U
Q01	a	Comparative bar charts - completing	0.9	1	90	0.9	0.67	0.89	0.95	0.97	0.98					0.9	0.21
Q01	b	Comparative bar charts - reading and using	1.69	2	85	1.69	1.19	1.63	1.77	1.85	1.9					1.69	0.36
Q01	c	Comparative bar charts - reading and using	1.13	2	56	1.13	0.31	0.75	1.17	1.47	1.71					1.13	0.05
Q01	d	Terms used to describe different types of data	0.43	1	43	0.43	0.08	0.23	0.4	0.6	0.75					0.43	
Q02	a	Drawing Tally charts	1.86	2	93	1.86	1.69	1.85	1.91	1.94	1.96					1.86	0.81
Q02	b	Reading Tally charts	0.96	1	96	0.96	0.91	0.97	0.98	0.99	0.99					0.96	0.44
Q02	c	Tally charts - estimates of probabilities	0.85	1	85	0.85	0.46	0.79	0.92	0.96	0.99					0.85	0.08
Q02	d	Interpreting Tally charts	0.73	1	73	0.73	0.41	0.65	0.74	0.85	0.9					0.73	0.1
Q02	e	Selecting appropriate types of average	0.03	1	3	0.03		0.01	0.01	0.04	0.11					0.03	
Q02	f	Justify the appropriate type of chart/diagram	0.66	1	66	0.66	0.28	0.52	0.69	0.81	0.88					0.66	0.03
Q02	g	Justify the appropriate type of chart/diagram	0.9	1	90	0.9	0.72	0.89	0.94	0.97	0.98					0.9	0.24
Q03	a	Experimental data / theoretical predictions to identify bias	0.58	1	58	0.58	0.18	0.41	0.59	0.75	0.87					0.58	0.02
Q03	b	Experimental data / theoretical predictions to identify bias	1.36	2	68	1.36	0.39	1.06	1.47	1.69	1.84					1.36	0.02
Q03	c	Experimental probability will tend towards theoretical probability	0.46	1	46	0.46	0.13	0.3	0.45	0.58	0.76					0.46	0.01
Q04	a	Sampling techniques to avoid bias	0.3	1	30	0.3	0.05	0.15	0.26	0.43	0.58					0.3	
Q04	b	Reading line graphs	1.47	2	74	1.47	0.53	1.22	1.6	1.8	1.88					1.47	0.02
Q04	c	Calculate mode	0.69	1	69	0.69	0.2	0.51	0.74	0.89	0.96					0.69	0.04
Q04	d	Assess appropriateness of a conclusion	0.5	1	50	0.5	0.07	0.28	0.49	0.69	0.82					0.5	0.01
Q05	a	Complete two-way tables	1.9	2	95	1.9	1.56	1.95	1.98	1.99	1.99					1.9	0.4
Q05	b i	Calculate estimates of probabilities	0.58	1	58	0.58	0.15	0.44	0.61	0.74	0.84					0.58	
Q05	b ii	Calculate estimates of probabilities	0.5	1	50	0.5	0.08	0.29	0.5	0.7	0.8					0.5	
Q05	b iii	Calculate estimates of probabilities	1.07	2	54	1.07	0.22	0.7	1.11	1.39	1.66					1.07	0.01
Q05	c	Reading two-way tables	1.43	2	72	1.43	0.5	1.17	1.55	1.76	1.85					1.43	0.02
Q06	a	Sources of data	0.95	1	95	0.95	0.79	0.93	0.98	0.99	0.99					0.95	0.45
Q06	b	Difference between primary and secondary data	0.65	1	65	0.65	0.14	0.47	0.68	0.84	0.95					0.65	0.01
Q06	c	Compare data sets using appropriate measure of central tendency	1.86	3	62	1.86	0.08	0.82	2.09	2.68	2.84					1.86	0.01
Q07	a	Quota sampling	0.05	1	5	0.05		0.02	0.03	0.07	0.16					0.05	
Q07	b	Population	0.03	1	3	0.03	0.01	0.01	0.02	0.04	0.09					0.03	
Q07	c	Sampling techniques to avoid bias	0.66	3	22	0.66	0.06	0.26	0.54	0.95	1.39					0.66	
Q08	a	Sources of data	0.6	2	30	0.6	0.09	0.3	0.59	0.86	1.01					0.6	
Q08	b	Importance of identifying and controlling extraneous variables	0.32	1	32	0.32	0.05	0.16	0.32	0.46	0.55					0.32	
Q09	a	Rates of change over time	0.27	1	27	0.27	0.03	0.1	0.25	0.38	0.54					0.27	
Q09	b	Rates of change over time	0.04	1	4	0.04			0.02	0.04	0.13					0.04	
Q09	c	Calculate crude birth rate	1.44	2	72	1.44	0.51	1.15	1.57	1.78	1.88					1.44	0.03
Q10	a i	Advantages of grouping numerical data into class intervals	0.35	1	35	0.35	0.06	0.15	0.3	0.5	0.67					0.35	
Q10	a ii	Disdvantages of grouping numerical data into class intervals	0.23	1	23	0.23	0.04	0.1	0.19	0.33	0.48					0.23	
Q10	b	Limitations of grouping numerical data into class intervals	0.32	1	32	0.32	0.03	0.15	0.3	0.46	0.6					0.32	
Q10	c	Identify outliers	0.35	1	35	0.35	0.01	0.08	0.29	0.55	0.76					0.35	
Q10	d i	Determine skewness	0.16	1	16	0.16	0.01	0.07	0.13	0.22	0.36					0.16	
Q10	d ii	Interpret a distribution of data in term of skewness	0	1	0						0.02						
Q11	a	Drawing Venn diagrams	2.29	5	46	2.29	0.42	1.22	2.18	3.15	4					2.29	0.04
Q11	b	Importance of reliability and validity	0.81	3	27	0.81	0.03	0.25	0.67	1.23	1.72					0.81	
Q12		Box plots	0.57	5	11	0.57		0.01	0.13	0.74	2.23					0.57	
Q13	a	Advantages of Cumulative frequency charts	0.05	1	5	0.05		0.01	0.03	0.07	0.17					0.05	
Q13	b	Calculate mode	0.35	1	35	0.35	0.04	0.19	0.34	0.47	0.62					0.35	0.01
Q13	c i	Reading CF charts	0.08	1	8	0.08		0.02	0.04	0.1	0.26					0.08	
Q13	c ii	Reading CF charts	0.55	2	28	0.55	0.05	0.19	0.45	0.77	1.23					0.55	
Q13	d	Reading CF charts	0.21	1	21	0.21	0.05	0.13	0.21	0.26	0.36					0.21	
Q13	e	Calculate interquartile range (IQR)	0.01	1	1	0.01				0.01	0.02					0.01	
Q14	a	Reading a Choropleth map	1.43	3	48	1.43	0.11	0.61	1.45	2.04	2.49					1.43	
Q14	b	Using a Choropleth map	1.17	2	59	1.17	0.14	0.61	1.25	1.65	1.83					1.17	0.01
Q14	c	Determine factors that may lead to bias	0.09	1	9	0.09		0.01	0.05	0.13	0.25					0.09	
			36.87	80		36.87	13.53	25.68	36.93	46.63	55.6	0	0	0	0	36.87	3.43

Foundation P2 Q1 (a/b)

Student A

1 The incomplete comparative bar chart shows the total number of medals won by three of the countries that took part in the 2014 and 2018 Winter Olympics.



(Source: *www.statista.com*)

The total number of medals won by Sweden in the 2018 Winter Olympics was 14

(*a*) Complete the comparative bar chart for Sweden.

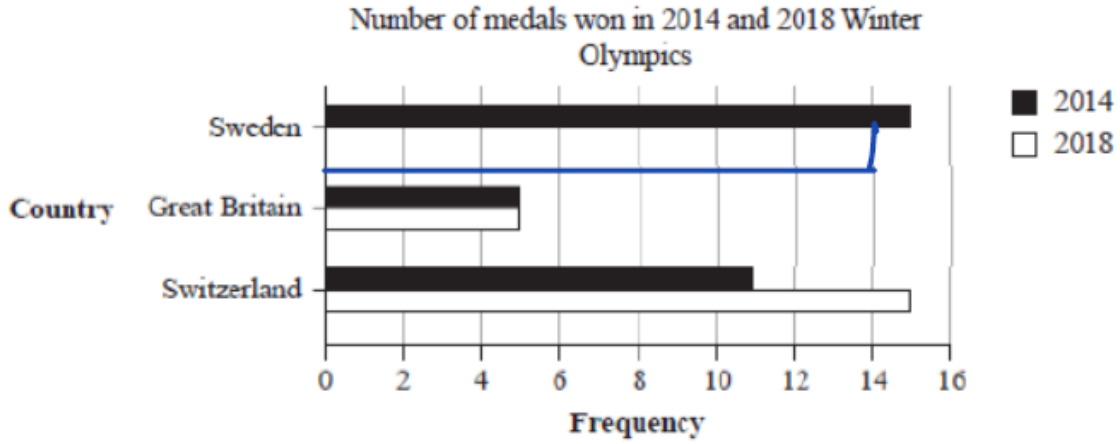
(1)

(*b*) Work out how many more medals were won by Sweden than Great Britain in the 2014 Winter Olympics. 15 5

10

1(a)	B1 bar drawn at length of 14 and not shaded.	B1 Ignore bar width and a bar not drawn with a ruler.	(1)
(b)	M1 15 -5	M1 15 $-k$ or $k-5$ OR 5 and 15 identified	(2)

1 The incomplete comparative bar chart shows the total number of medals won by three of the countries that took part in the 2014 and 2018 Winter Olympics.



(Source: *www.statista.com*)

The total number of medals won by Sweden in the 2018 Winter Olympics was 14

(*a*) Complete the comparative bar chart for Sweden.

(1)

(*b*) Work out how many more medals were won by Sweden than Great Britain in the 2014 Winter Olympics.

5 + 5 = 10 14 + 15 = 29

29 - 10

19

Foundation P2 Q1 (a/b)

Student B

1(a)	B1 bar drawn at length of 14 and not shaded.	B1 Ignore bar width and a bar not drawn with a ruler.	(1)
(b)	M1 15 -5	M1 15 $-$ k or k $-$ 5 OR 5 and 15 identified	(2)

(c) Compare the total number of medals won by Sweden, Great Britain and Switzerland in the 2014 Winter Olympics.

Sweden won the most with 29 medals then Switzerland then Great Britain

(2)

Thomas says that the data displayed in the comparative bar chart is quantitative data.

(d) Explain what is meant by quantitative data.

Quantitative is things you can count.

(1)

(Total for Question 1 is 6 marks)

(c)	B2 for any two from ; • Sweden won the most/Sweden won (4) more than Switzerland/ Sweden won (10) more than Great Britain • Great Britain won the least • Switzerland had the second most/Switzerland won (6) more than Great Britain (B1 for one correct comparison)	Accept converse statements If figures are given they need to be correct B0 for comparing the total number of medals won in 2014 and 2018 for each country.	(2)
(d)	B1 Data that is numeric or numbers.	Do not accept answers referencing integer/quantity/data that you count	(1)

Foundation P2 Q1 (c/d) Student A

(c) Compare the total number of medals won by Sweden, Great Britain and Switzerland in the 2014 Winter Olympics.

1st = Sweden, 2nd = Switzerland
3rd = GB.

(2)

Thomas says that the data displayed in the comparative bar chart is quantitative data.

(d) Explain what is meant by quantitative data.

They are numbers.

(1)

(Total for Question 1 is 6 marks)

(c)	B2 for any two from ; • Sweden won the most/Sweden won (4) more than Switzerland/ Sweden won (10) more than Great Britain • Great Britain won the least • Switzerland had the second most/Switzerland won (6) more than Great Britain (B1 for one correct comparison)	Accept converse statements If figures are given they need to be correct B0 for comparing the total number of medals won in 2014 and 2018 for each country.	(2)
(d)	B1 Data that is numeric or numbers.	Do not accept answers referencing integer/quantity/data that you count	(1)

Foundation P2 Q1 (c/d) Student B

Foundation Paper 2 Q1

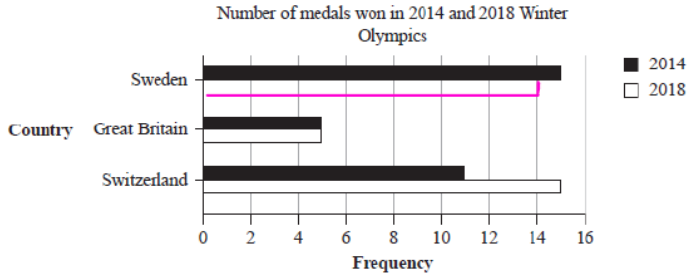
WAGOLL

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 The incomplete comparative bar chart shows the total number of medals won by three of the countries that took part in the 2014 and 2018 Winter Olympics.



(Source: *www.statista.com*)

The total number of medals won by Sweden in the 2018 Winter Olympics was 14

(*a*) Complete the comparative bar chart for Sweden.

(1)

(*b*) Work out how many more medals were won by Sweden than Great Britain in the 2014 Winter Olympics.

15 5

15 - 5 = 10

10

(2)

(*c*) Compare the total number of medals won by Sweden, Great Britain and Switzerland in the 2014 Winter Olympics.

15 7 11

Sweden won the most medals, Switzerland had second most and GB won the least

(2)

Thomas says that the data displayed in the comparative bar chart is quantitative data.

(*d*) Explain what is meant by quantitative data.

Data that is numerical ie. can be counted.

(1)

(Total for Question 1 is 6 marks)

Norbert asked each of the students in his class to name their favourite fruit from Apple, Banana, Orange or Pear.
The results are shown below.

Banana	Orange	Apple	Banana	Pear	Apple
Apple	Banana	Orange	Pear	Apple	Banana	Apple
Apple	Apple	Orange	Apple	Pear	Banana	Banana

(*a*) Fill in the tally chart for this information **and** complete the frequency column.

Fruit	Tally	Frequency
Apple	𝍸 III	8 ✓
Banana	𝍸 I	6
Orange	II	2 ✗
Pear	II	2

(2)

(*b*) How many students are in the class?

18 ✓

(1)

One of the students is chosen at random.

(*c*) Find the probability that this student's favourite fruit is Orange.

$\frac{2}{18}$ ✓

(1)

(*d*) Compare the number of students whose favourite fruit is Apple to the number of students whose favourite fruit is Pear.

8 2 ✗

4 x 2 = 8

(1)

2(a)	M1 A1 Frequencies 8,6,3,3	M1 At least one tally or frequency correct A1 all frequencies correct (tallies not required and incorrect tallies can be ignored if frequencies are correct)	(2)
(b)	B1ft 20	B1ft from their frequency table total	(1)
(c)	B1ft $\frac{"3"}{"20"}$ oe	B1ft their numerator for their frequency for orange from the table and ft their answer to part (b) for their denominator	(1)
(d)	B1ft More students preferred apples	Accept for example: B1ft 5 more students preferred apple or the difference is 5. If figures are stated they need to be correct or ft from their tally chart. B0 for just stating figures	(1)

Foundation P2 Q2 (a/b/c/d)

Student A

Norbert asked each of the students in his class to name their favourite fruit from Apple, Banana, Orange or Pear.
The results are shown below.

Banana	Orange	Apple	Banana	Pear	Apple
Apple	Banana	Orange	Pear	Apple	Banana	Apple
Apple	Apple	Orange	Apple	Pear	Banana	Banana

(a) Fill in the tally chart for this information **and** complete the frequency column.

Fruit	Tally	Frequency
Apple		8 ✓
Banana		6 ✓
Orange		3
Pear		3

(2)

(b) How many students are in the class?

20 ✓

.....

(1)

One of the students is chosen at random.

(c) Find the probability that this student's favourite fruit is Orange.

$\frac{3}{20}$ ✓

.....

(1)

(d) Compare the number of students whose favourite fruit is Apple to the number of students whose favourite fruit is Pear. 3 8 ✓

.....

5 more liked Apple.

.....

(1)

2(a)	M1 A1 Frequencies 8,6,3,3	M1 At least one tally or frequency correct A1 all frequencies correct (tallies not required and incorrect tallies can be ignored if frequencies are correct)	(2)
(b)	B1ft 20	B1ft from their frequency table total	(1)
(c)	B1ft $\frac{\text{"3"}}{\text{"20"}}$ oe	B1ft their numerator for their frequency for orange from the table and ft their answer to part (b) for their denominator	(1)
(d)	B1ft More students preferred apples	Accept for example; B1ft 5 more students preferred apple or the difference is 5. If figures are stated they need to be correct or ft from their tally chart. B0 for just stating figures	(1)

Foundation P2 Q2 (a/b/c/d)

Student B

Foundation P2 Q2 (e/f/g)

Student A

Norbert decides to find the favourite fruit that is the mode.

(e) Explain why the mode is an appropriate average for Norbert to find for this type of data.

mode is easy to find. ✗

.....

.....

(1)

(f) Give one advantage of the tally chart over the raw data.

it is clearer to read ✓

.....

.....

(1)

Norbert wants to draw a diagram to represent his results.

(g) Circle the type of diagram from the list below that is most suitable for him to draw.

Scatter diagram | Bar chart ✓ | Line graph | Time series

(1)

(Total for Question 2 is 8 marks)

(e)	B1 for one of • The data is non numeric • Qualitative (data) • (Data is) not numbers • Not possible to find the mean or median.	B1 for a correct reason. Do not allow e.g. 'easy to find', 'gives the most popular' which are both B0	(1)
(f)	B1 for one of • easier to analyse data • Data is more organised • Easier to read off the most/least popular/read off frequencies/see the differences	B1 for any equivalent statement about the advantage of a tally chart Accept: clearer to see results B0 for it is more reliable or more accurate.	(1)

(g)	B1 bar chart	B1 for selecting correct option only.	(1)

Norbert decides to find the favourite fruit that is the mode.

(e) Explain why the mode is an appropriate average for Norbert to find for this type of data.

you can't find the mean, or median or range.

(1)

(f) Give one advantage of the tally chart over the raw data.

Easier to analysis.

(1)

Norbert wants to draw a diagram to represent his results.

(g) Circle the type of diagram from the list below that is most suitable for him to draw.

Scatter diagram | Bar chart | Line graph | Time series

(1)

(Total for Question 2 is 8 marks)

Foundation P2 Q2 (e/f/g)

Student B

(e)	B1 for one of • The data is non numeric • Qualitative (data) • (Data is) not numbers • Not possible to find the mean or median.	B1 for a correct reason. Do not allow e.g. 'easy to find', 'gives the most popular' which are both B0	(1)
(f)	B1 for one of • easier to analyse data • Data is more organised • Easier to read off the most/least popular/read off frequencies/see the differences	B1 for any equivalent statement about the advantage of a tally chart Accept: clearer to see results B0 for it is more reliable or more accurate.	(1)

(g)	B1 bar chart	B1 for selecting correct option only.	(1)

Foundation Paper 2 Q2

WAGOLL

2 Norbert asked each of the students in his class to name their favourite fruit from Apple, Banana, Orange or Pear.
The results are shown below.

Banana Orange Apple Banana Pear Apple
Apple Banana Orange Pear Apple Banana Apple
Apple Apple Orange Apple Pear Banana Banana

(*a*) Fill in the tally chart for this information **and** complete the frequency column.

Fruit	Tally	Frequency
Apple	𝍸 III	8
Banana	𝍸 I	6
Orange	III	3
Pear	III	3

(2)

(*b*) How many students are in the class?

8 + 6 + 3 + 3

20

(1)

One of the students is chosen at random.

(*c*) Find the probability that this student's favourite fruit is Orange.

$\frac{3}{20}$

(1)

(*d*) Compare the number of students whose favourite fruit is Apple to the number of students whose favourite fruit is Pear. 8 3

More students preferred Apple

(1)

Norbert decides to find the favourite fruit that is the mode.

(*e*) Explain why the mode is an appropriate average for Norbert to find for this type of data.

Its not possible to find the mean or median because the data is qualitative.

(1)

(*f*) Give one advantage of the tally chart over the raw data.

eg. Its easier to read the frequencies of each type of fruit

(1)

Norbert wants to draw a diagram to represent his results.

(*g*) Circle the type of diagram from the list below that is most suitable for him to draw.

Scatter diagram **Bar chart** Line graph Time series

(1)

(Total for Question 2 is 8 marks)

Sam used the internet to collect the times, in minutes, it took for 50 cyclists to compete in a hill climb competition. He used a group frequency table to record the results he collected.

(*a*) (i) Give one advantage of using grouped data rather than raw data.

It's quicker ✗

(1)

(ii) Give one disadvantage of using grouped data rather than raw data.

estimates can only be calculated ✓

(1)

Sam used this grouped frequency table to show the results for the hill climb.

Time (t minutes)	Frequency
$11 \leq t < 12$	2
$12 \leq t < 13$	25
$13 \leq t < 14$	15
$14 \leq t < 15$	4
$15 \leq t < 16$	1
$16 \leq t < 17$	1
$17 \leq t < 18$	1

(Source: *cyclin*

Before Sam collected the data he did not know what the longest time would be.
The longest time in the hill climb was 28.3 minutes.

(*b*) Explain why this table cannot be used to show the data for all 50 riders.

It only contains data for 49 riders. ✗

(1)

10(a) (i)	B1 Any from… • Grouping data can help to spot patterns in the data • Makes it easier to process large amounts of data • Easy to compare different groups • The data is easier to read • The data is easier to represent on graphs	Accept any other equivalent answer B0 for more accurate/easier/quicker	(1)
(a)(ii)	B1 Any from… • Some detail is lost using grouped data/raw data retains the actual data • We can only calculate estimates of statistical values /statistics can be calculated exactly from raw data	B0 for not reliable	(1)
(b)	B1 Couldn't record the longest time in the table OR didn't allow him to record any data past 18 minutes.	B1 Some reference to been unable to record the longest time Reference to the table only containing data for 49 riders is B0. Do not accept it is an outlier.	(1)
(c)	B1 Outlier	B1 accept anomaly or it is far	(1)

Foundation P1 Q10 (a/b)

Student A

10 Sam used the internet to collect the times, in minutes, it took for 50 cyclists to compete in a hill climb competition. He used a group frequency table to record the results he collected.

(*a*) (i) Give one advantage of using grouped data rather than raw data.

Easier to use. ✗

(1)

(ii) Give one disadvantage of using grouped data rather than raw data.

Detail is lost using grouped data ✓

(1)

Sam used this grouped frequency table to show the results for the hill climb.

Time (t minutes)	Frequency
$11 \leq t < 12$	2
$12 \leq t < 13$	25
$13 \leq t < 14$	15
$14 \leq t < 15$	4
$15 \leq t < 16$	1
$16 \leq t < 17$	1
$17 \leq t < 18$	1

(Source: *cyclin*

Before Sam collected the data he did not know what the longest time would be.
The longest time in the hill climb was 28.3 minutes.

(*b*) Explain why this table cannot be used to show the data for all 50 riders.

only goes up to 18. ✗

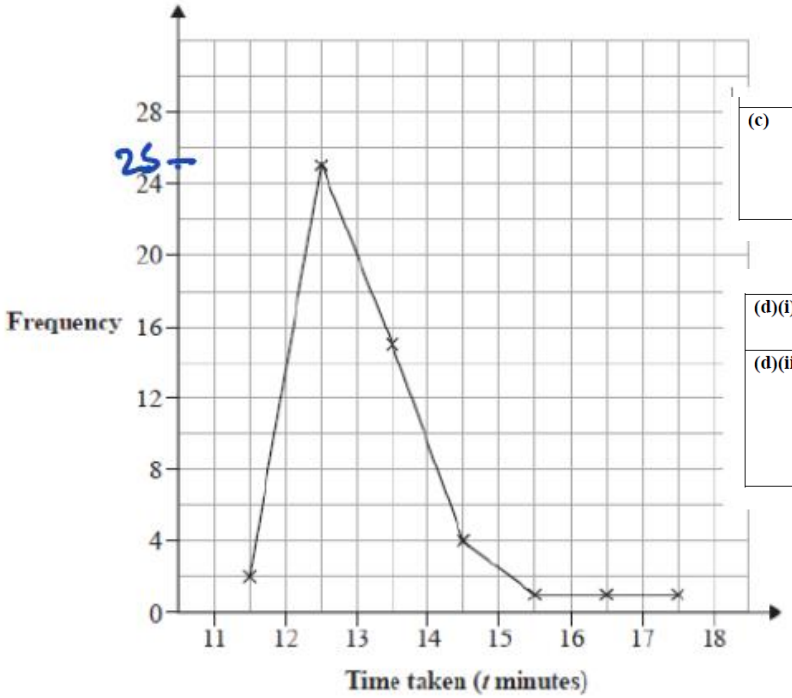
(1)

10(a) (i)	B1 Any from… • Grouping data can help to spot patterns in the data • Makes it easier to process large amounts of data • Easy to compare different groups • The data is easier to read • The data is easier to represent on graphs	Accept any other equivalent answer B0 for more accurate/easier/quicker	(1)
(a)(ii)	B1 Any from… • Some detail is lost using grouped data/raw data retains the actual data • We can only calculate estimates of statistical values /statistics can be calculated exactly from raw data	B0 for not reliable	(1)
(b)	B1 Couldn't record the longest time in the table OR didn't allow him to record any data past 18 minutes.	B1 Some reference to been unable to record the longest time Reference to the table only containing data for 49 riders is B0. Do not accept it is an outlier.	(1)
(c)	B1 Outlier	B1 accept anomaly or it is far	(1)

Foundation P1 Q10 (a/b)

Student B

Sam drew this frequency polygon for the hill climb results.



(c)	B1 Outlier	B1 accept anomaly or it is far from the rest of the data. Allow any equivalent description Do not accept outliner for outlier.	(1)

(d)(i)	B1 Positive (skew)		(1)
(d)(ii)	B1ft e.g. More than half of the cyclists took less than the mean time cycling up the hill Or B1 Cyclists who take longer than the median time are more varied in their times than those who take less time than the median time	B1ft for correct interpretation of the skewness in part (i) Any reference to the spread of values below/above the mean is B0.	(1)

Sam decided not to include the value of 28.3 minutes on his frequency polygon.

(*c*) Suggest a reason why Sam's decision might be appropriate.

the graph only goes to 18 ✗

(1)

(*d*) (i) Describe the skew of the distribution.

Negative ✗

(1)

(ii) Interpret the skew of the distribution in context.

25 riders. ✗

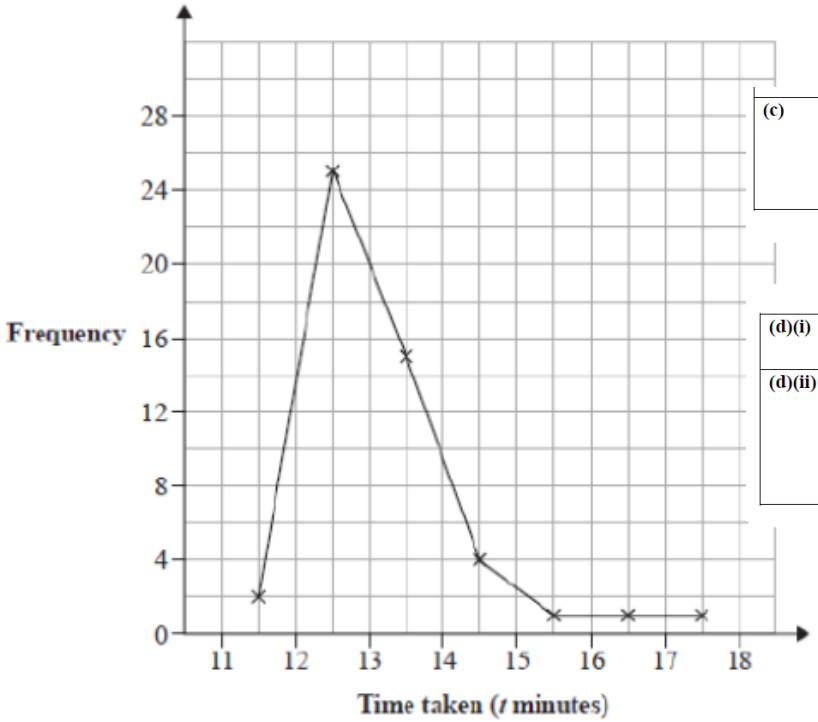
(1)

(Total for Question 10 is 6 marks)

Foundation P1 Q10 (c/d)

Student A

Sam drew this frequency polygon for the hill climb results.



Sam decided not to include the value of 28.3 minutes on his frequency polygon.

(c) Suggest a reason why Sam's decision might be appropriate.

its anomoly. ✓

.....

.....

(1)

(d) (i) Describe the skew of the distribution.

positive. ✓

.....

(1)

(ii) Interpret the skew of the distribution in context.

its high at the start then gets lowe ✗

.....

.....

(1)

(Total for Question 10 is 6 marks)

(c)	B1 Outlier	B1 accept anomaly or it is far from the rest of the data. Allow any equivalent description Do not accept outliner for outlier.	(1)

(d)(i)	B1 Positive (skew)		(1)
(d)(ii)	B1ft e.g. More than half of the cyclists took less than the mean time cycling up the hill Or B1 Cyclists who take longer than the median time are more varied in their times than those who take less time than the median time	B1ft for correct interpretation of the skewness in part (i) Any reference to the spread of values below/above the mean is B0.	(1)

Foundation P1 Q10 (c/d)

Student B

Foundation Paper 2 Q10

WAGOLL

10 Sam used the internet to collect the times, in minutes, it took for 50 cyclists to compete in a hill climb competition. He used a group frequency table to record the results he collected.

(a) (i) Give one advantage of using grouped data rather than raw data.

e.g. grouping data makes it easy to compare different groups.

(1)

(ii) Give one disadvantage of using grouped data rather than raw data.

e.g. we can only calculate 'estimates' for some statistics.

(1)

Sam used this grouped frequency table to show the results for the hill climb.

Time (t minutes)	Frequency
$11 \leq t < 12$	2
$12 \leq t < 13$	25
$13 \leq t < 14$	15
$14 \leq t < 15$	4
$15 \leq t < 16$	1
$16 \leq t < 17$	1
$17 \leq t < 18$	1

(Source: *cyclinguphill.com*)

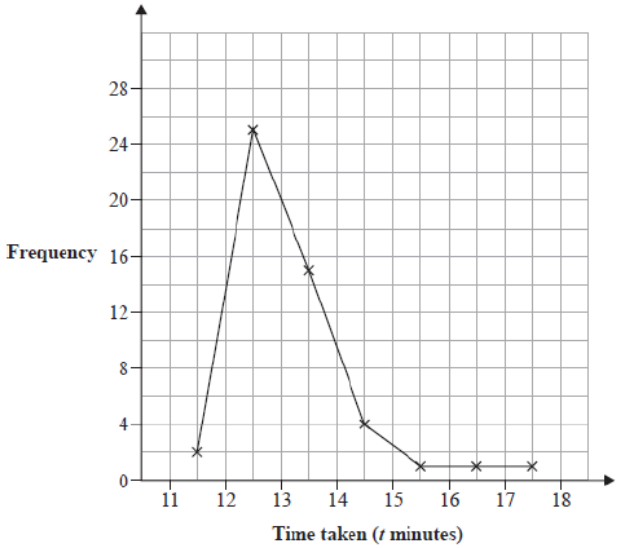
Before Sam collected the data he did not know what the longest time would be.
The longest time in the hill climb was 28.3 minutes.

(b) Explain why this table cannot be used to show the data for all 50 riders.

The table only goes up to a time of 18 minutes.

(1)

Sam drew this frequency polygon for the hill climb results.



Sam decided not to include the value of 28.3 minutes on his frequency polygon.

(c) Suggest a reason why Sam's decision might be appropriate.

28.3 minutes may be an outlier.

(1)

(d) (i) Describe the skew of the distribution.

Positive

(1)

(ii) Interpret the skew of the distribution in context.

More than half the riders took less than the mean time cycling up the hill.

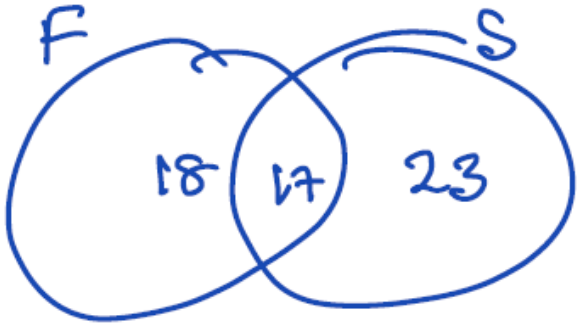
(1)

(Total for Question 10 is 6 marks)

Grace asked a sample of 60 people in her town if they had ever visited France or Spain.

17 people visited both France and Spain
23 people visited Spain only
33 people visited France

(*a*) Draw a Venn diagram to represent this information.



(5)

Grace says

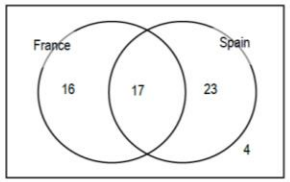
- more than half of the people in her sample have visited France
- therefore more than half of the people in her town have visited France

(*b*) Discuss the validity of each of Grace's comments.

18 + 17 = 35 35 is more than 30.

(3)

(Total for Question 11 is 8 marks)

1(a)	 M1 two intersecting circles drawn A1 17 in centre and 23 in Spain M1 33 – 17 or 16 M1 60 – ('16' + 17 + 23) or 4 in the correct region A1 fully correct with labels	Accept e.g. F,S or any other unambiguous indication for labels. Do not accept three intersecting circles. Their ft '16' must be ≤ 20	(5)
b)	B1 e.g. • The statement is valid for her sample since 33/60 is more than ½ • The statement is valid as 0.55 is greater than 0.5 • The statement is valid as 33 is more than half of 60 B2 e.g. Not valid/we cannot be sure since we don't know if the sample was representative. (B1 e.g. 'Not Valid/We cannot be sure' with an attempt at a reason OR sample may not be representative/sample is too small/do not know the population size of the town with no conclusion/incorrect conclusion)	B1 for first statement valid with supporting reason Allow: Valid, since 33 is more than half. B2 for second statement may be valid with supporting reason (B1 for not valid/unable to determine the validity of the second statement with an attempt at a reason OR for a correct comment on issues with the sample with no conclusion/incorrect conclusion)	(3)

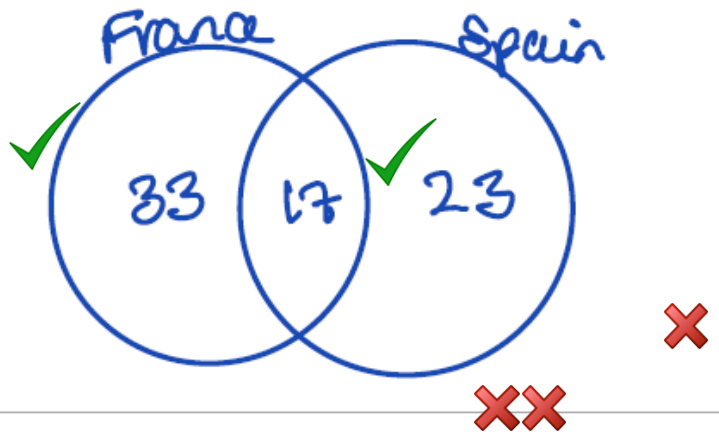
COMMON ITEM
Foundation P2 Q11 (and Higher P2 Q1)

Student A

Grace asked a sample of 60 people in her town if they had ever visited France or Spain.

17 people visited both France and Spain
23 people visited Spain only
33 people visited France

(*a*) Draw a Venn diagram to represent this information.



(5)

Grace says

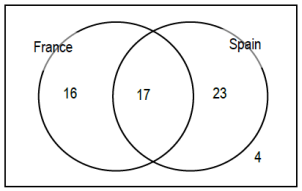
- more than half of the people in her sample have visited France
- therefore more than half of the people in her town have visited France

(*b*) Discuss the validity of each of Grace's comments.

33+17 = 50
50 > 30
hes statement is valid - ✗✗

(3)

(Total for Question 11 is 8 marks)

l(a)	 M1 two intersecting circles drawn A1 17 in centre and 23 in Spain M1 33 – 17 or 16 M1 60 – ('16' + 17 + 23) or 4 in the correct region A1 fully correct with labels	Accept e.g. F,S or any other unambiguous indication for labels. Do not accept three intersecting circles. Their ft '16' must be ≤ 20	(5)
)	B1 e.g. • The statement is valid for her sample since 33/60 is more than ½ • The statement is valid as 0.55 is greater than 0.5 • The statement is valid as 33 is more than half of 60 B2 e.g. Not valid/we cannot be sure since we don't know if the sample was representative. (B1 e.g. 'Not Valid/We cannot be sure' with an attempt at a reason OR sample may not be representative/sample is too small/do not know the population size of the town with no conclusion/incorrect conclusion)	B1 for first statement valid with supporting reason Allow: Valid, since 33 is more than half. B2 for second statement may be valid with supporting reason (B1 for not valid/unable to determine the validity of the second statement with an attempt at a reason OR for a correct comment on issues with the sample with no conclusion/incorrect conclusion)	(3)

COMMON ITEM
Foundation P2 Q11 (and Higher P2 Q1)

Student B

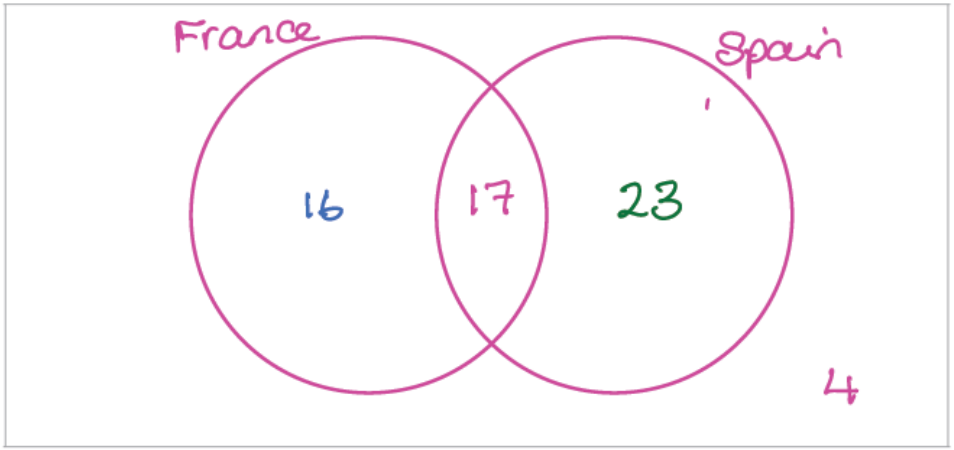
COMMON ITEM
Foundation P2 Q11 (and Higher P2 Q1)

WAGOLL

Grace asked a sample of 60 people in her town if they had ever visited France or Spain.

17 people visited both France and Spain ✓
23 people visited Spain only ✓
33 people visited France 33 − 17 = 16

(*a*) Draw a Venn diagram to represent this information. 60 − (16 + 17 + 23) = 4



(5)

Grace says

- more than half of the people in her sample have visited France
- therefore more than half of the people in her town have visited France

(*b*) Discuss the validity of each of Grace's comments.

The number who visited France = 16 + 17 = 33
which is greater than half of 60 (30)
The first statement is valid.
But the second statement is not valid because
we don't know if the sample is representative.

(3)

Higher Paper 2

H P2

Q	Part	Skill tested	Score	Max score	Percent	ALL	1	2	Edexcel averages: 3	4	5	6	7	8	9	ALL	U
Q01	a	Drawing Venn diagrams	4.26	5	85	4.26			2.74	3.43	4.02	4.46	4.67	4.87	4.94	4.26	2.02
Q01	b	Importance of reliability and validity	1.27	3	42	1.27			0.35	0.65	0.97	1.26	1.48	1.85	2.33	1.27	0.24
Q02		Box plots	2.83	5	57	2.83			0.41	1.18	2.23	3.07	3.56	4.02	4.48	2.83	0.12
Q03	a	Advantages of Cumulative frequency charts	0.38	1	38	0.38			0.04	0.08	0.16	0.34	0.52	0.72	0.9	0.38	0.03
Q03	b	Calculate mode	0.71	1	71	0.71			0.35	0.53	0.65	0.73	0.79	0.87	0.92	0.71	0.18
Q03	c i	Reading CF charts	0.38	1	38	0.38			0.04	0.12	0.22	0.36	0.5	0.65	0.83	0.38	0.01
Q03	c ii	Reading CF charts	1.39	2	70	1.39			0.54	0.88	1.17	1.47	1.66	1.8	1.91	1.39	0.24
Q03	d	Reading CF charts	0.37	1	37	0.37			0.24	0.26	0.32	0.35	0.4	0.45	0.6	0.37	0.14
Q03	e	Calculate interquartile range (IQR)	0.05	1	5	0.05			0.01	0.01	0.03	0.03	0.04	0.07	0.18	0.05	0.01
Q04	a	Difference between population, sample frame and sample	0.39	1	39	0.39			0.31	0.27	0.27	0.29	0.41	0.57	0.79	0.39	0.4
Q04	b ii	Random, systematic, and quota sampling	0.97	2	49	0.97			0.2	0.52	0.77	1.02	1.19	1.34	1.45	0.97	0.07
Q04	b ii	Random, systematic, and quota sampling	0.9	2	45	0.9			0.1	0.37	0.59	0.89	1.17	1.44	1.62	0.9	0.04
Q05		Compare experimental data with theoretical predictions to identify bias	2.31	4	58	2.31			0.79	1.42	2	2.35	2.7	3.02	3.44	2.31	0.32
Q06	a	Frequency polygons	0.78	1	78	0.78			0.42	0.56	0.7	0.83	0.9	0.95	0.99	0.78	0.16
Q06	b	Frequency polygons	1.85	2	93	1.85			1.22	1.63	1.86	1.93	1.96	1.98	1.99	1.85	0.58
Q06	c	Use collected data to calculate estimates of probabilities	0.68	1	68	0.68			0.2	0.43	0.6	0.73	0.8	0.87	0.93	0.68	0.05
Q06	d	Determine skewness	0.53	2	27	0.53			0.01	0.06	0.13	0.35	0.74	1.17	1.68	0.53	0.01
Q06	e	Calculate interpercentile range and interdecile range	0.75	1	75	0.75			0.2	0.46	0.68	0.81	0.89	0.94	0.97	0.75	0.09
Q06	f	Calculate range, quartiles, interquartile range (IQR), percentiles	0.52	1	52	0.52			0.07	0.15	0.36	0.55	0.71	0.81	0.9	0.52	
Q07	a	Rates of change over time	1.89	2	95	1.89			1.71	1.82	1.88	1.91	1.94	1.95	1.98	1.89	1.15
Q07	b	Rates of change over time	0.86	2	43	0.86			0.15	0.37	0.6	0.88	1.06	1.29	1.62	0.86	0.06
Q08	a	Importance of identifying and controlling extraneous variables	0.51	2	26	0.51			0.1	0.21	0.32	0.42	0.59	0.86	1.23	0.51	0.04
Q08	b	Calculate weighted mean	1.72	3	57	1.72			0.2	0.46	0.98	1.76	2.55	2.85	2.95	1.72	0.06
Q09	a	Comparative pie chart	0.7	2	35	0.7			0.15	0.38	0.61	0.75	0.85	0.89	1.11	0.7	0.03
Q09	b	Comparative pie chart	0.64	3	21	0.64			0.01	0.03	0.09	0.3	0.79	1.62	2.47	0.64	
Q10		Use action and warning lines	2.43	5	49	2.43			0.54	1.18	1.75	2.39	3.12	3.65	4.15	2.43	0.18
Q11	a	Use means and standard deviation	0.77	1	77	0.77			0.17	0.38	0.68	0.88	0.96	0.99	1	0.77	0.06
Q11	b	Use means and standard deviation	1.75	3	58	1.75			0.24	0.42	0.95	1.85	2.63	2.88	2.98	1.75	0.1
Q11	c i	Appropriateness of using sample means	0.11	2	6	0.11			0.01	0.02	0.04	0.06	0.09	0.2	0.53	0.11	0.01
Q11	c ii	Appropriateness of using sample means	0.14	2	7	0.14			0.01	0.01	0.04	0.08	0.14	0.25	0.66	0.14	
Q12	a	Characteristics of a binomial distribution	1.96	6	33	1.96			0.05	0.2	0.53	1.28	2.86	4.51	5.5	1.96	0.01
Q12	b	Appropriateness of using a binomial distribution	1	3	33	1			0.01	0.03	0.2	0.65	1.53	2.39	2.86	1	0.01
Q12	c	Characteristics of a binomial distribution	0.21	1	21	0.21					0.03	0.11	0.3	0.53	0.73	0.21	
Q13	a	Sources of secondary data should be acknowledged	0.18	1	18	0.18			0.06	0.1	0.14	0.17	0.21	0.24	0.34	0.18	0.03
Q13	b	Interpret PMCC	0.38	1	38	0.38			0.04	0.13	0.27	0.38	0.47	0.57	0.72	0.38	0.01
Q13	c i	Distinction between Spearman's rank correlation coefficient and PMCC	0.37	2	19	0.37			0.03	0.05	0.08	0.17	0.41	0.9	1.43	0.37	0.01
Q13	d	Interpret Pearson's product moment correlation coefficient (PMCC)	0.42	2	21	0.42			0.01	0.02	0.1	0.24	0.52	0.95	1.52	0.42	
			37.36	80		37.36	0	0	11.73	18.82	27.02	36.1	46.11	55.91	65.63	37.36	6.47

The table shows information about the death rate in England and Wales in 2000

Year	Total population	Number of deaths
2000	52 140 181	537 877

(Source: w

(a) Using the formula below, calculate the crude death rate for England and Wal in 2000

$$\text{crude death rate} = \frac{\text{number of deaths} \times 1000}{\text{total population}}$$

537877 x1000 ✓
52140181

10. ✗

(2)

In 2000, the crude birth rate in England and Wales was 11.6
In 2019, the crude birth rate in England and Wales was 10.8

A newspaper prints the following headline for an article.

> 36 000 more births in England and Wales in 2019 than in 2000

(b) Discuss whether or not this headline could be true given that the crude birth rate in 2019 is less than the crude birth rate in 2000

it could be true because we ✓
don't know the actual ✗
births.

(2)

(Total for Question 7 is 4 marks)

number			
7(a)	M1 $\frac{537\,877 \times 1000}{52\,140\,181}$ A1 10.315….	M1 correct calculation which may be seen in stages Do not allow a misread of 100 for 1000. A1 answer in the range 10.3 to 10.4	(2)
(b)	B2 e.g. 'Can be true as the population has likely increased' (B1 for can be true with an incomplete reason e.g. Can be true because the population has changed OR For the correct reason with no or incorrect conclusion)	B2 Can be true with correct supporting reason (B1 Can be true with incomplete reasoning)	(2)

Higher P2 Q7

Student A

The table shows information about the death rate in England and Wales in 2000

Year	Total population	Number of deaths
2000	52 140 181	537 877

(Source: ww

(*a*) Using the formula below, calculate the crude death rate for England and Wale in 2000

$$\text{crude death rate} = \frac{\text{number of deaths} \times 1000}{\text{total population}}$$

537877 x 100 / 52140181

1.031

(2)

In 2000, the crude birth rate in England and Wales was 11.6
In 2019, the crude birth rate in England and Wales was 10.8

A newspaper prints the following headline for an article.

> 36 000 more births in England and Wales in 2019 than in 2000

(*b*) Discuss whether or not this headline could be true given that the crude birth rate in 2019 is less than the crude birth rate in 2000

No, 11.6 is less than 10.8

(2)

(Total for Question 7 is 4 marks)

number			
7(a)	M1 $\frac{537\,877 \times 1000}{52\,140\,181}$ A1 10.315….	M1 correct calculation which may be seen in stages Do not allow a misread of 100 for 1000. A1 answer in the range 10.3 to 10.4	(2)
(b)	B2 e.g. 'Can be true as the population has likely increased' (B1 for can be true with an incomplete reason e.g. Can be true because the population has changed OR For the correct reason with no or incorrect conclusion)	B2 Can be true with correct supporting reason (B1 Can be true with incomplete reasoning)	(2)

Higher P2 Q7

Student B

Higher Paper 2 Q7

WAGOLL

7 The table shows information about the death rate in England and Wales in 2000

Year	Total population	Number of deaths
2000	52 140 181	537 877

(Source: www.ons.gov.uk)

(a) Using the formula below, calculate the crude death rate for England and Wales in 2000

$$\text{crude death rate} = \frac{\text{number of deaths} \times 1000}{\text{total population}}$$

$\frac{537877 \times 1000}{52140181} = 10.315..$

10.3

(2)

In 2000, the crude birth rate in England and Wales was 11.6
In 2019, the crude birth rate in England and Wales was 10.8

A newspaper prints the following headline for an article.

> 36 000 more births in England and Wales in 2019 than in 2000

(b) Discuss whether or not this headline could be true given that the crude birth rate in 2019 is less than the crude birth rate in 2000

This could be true as the population is likely to have increased.

(2)

(Total for Question 7 is 4 marks)

Higher P2 Q11 (a/b)

Student A

The heights of a group of seven-year-old boys have a mean of 121.7 cm and a standard deviation of 5.3 cm.

(Source: *www.who.int/growthref/who2007_height_for_age/en/*)

Daniel is a seven-year-old boy with a standardised score for height of 0

(*a*) Write down Daniel's height.

121·7 cm

(1)

Syed and Timur are both seven-year-old boys.

Syed's standardised score for height is 1.4
Timur's standardised score for height is –1.6

Syed is taller than Timur.

(*b*) How much taller?

121·7 + 5·3 × 1·4 = 129·12

121·7 + 5·3 × −1·6 = −203·2

?

..... cm

(3)

number				
11(a)	B1 121.7			(1)
(b)	M2 $5.3\times(1.4-(-1.6))$	M1 $121.7+5.3\times1.4$ M1 $121.7+5.3\times(-1.6)$	For method 1 allow M1 for $1.4-(-1.6)$ or 3 provided there is no evidence of incorrect working.	(3)
	A1 15.9	A1 15.9		

The heights of a group of seven-year-old boys have a mean of 121.7 cm and a standard deviation of 5.3 cm.

(Source: *www.who.int/growthref/who2007_height_for_age/en/*)

Daniel is a seven-year-old boy with a standardised score for height of 0

(*a*) Write down Daniel's height.

121·7 ..... cm

(1)

Higher P2 Q11 (a/b)

Student B

Syed and Timur are both seven-year-old boys.

Syed's standardised score for height is 1.4
Timur's standardised score for height is –1.6

Syed is taller than Timur.

(*b*) How much taller?

1·4 - -1·6 = 3

3 ..... cm

(3)

number				
11(a)	B1 121.7			(1)
(b)	M2 $5.3\times(1.4-(-1.6))$	M1 $121.7+5.3\times1.4$ M1 $121.7+5.3\times(-1.6)$	For method 1 allow M1 for $1.4-(-1.6)$ or 3 provided there is no evidence of incorrect working.	(3)
	A1 15.9	A1 15.9		

Tarik takes a sample of 4 boys from the group of seven-year-old boys.
He wants to calculate the standardised score for the **sample mean** of their heights.
He uses 121.7 cm for the mean and 5.3 cm for the standard deviation.

(*c*) Discuss whether or not it is appropriate to use

(i) 121.7 cm as the mean in the calculation of the standardised score,

sample mean will be approximately the same as the population mean.

(2)

(ii) 5.3 cm as the standard deviation in the calculation of the standardised score.

(2)

(Total for Question 11 is 8 marks)

Higher P2 Q11 (c)

Student A

(c)(i)	B2 Appropriate since the sample mean will be approximately the same as the population mean (B1 for inappropriate/no conclusion since the sample mean will be approximately the same as the population mean)	B2 for appropriate and correct supporting reason Accept equivalent wording for sample and/or population. (B1 for correct reason with no or incorrect conclusion)
(c)(ii)	B2 Not appropriate since the spread of the sample mean is smaller than individual values of the population (B1 for appropriate/no conclusion, because the spread of the sample mean is smaller than individual values of the population OR B1 for not appropriate because the spread of the sample mean is different to/not the same as the spread of the individual values of the population)	B2 for not appropriate and correct supporting reason (B1 for correct conclusion with incomplete reason, OR for correct reason with no or incorrect conclusion)

Higher Paper 2 Q11

WAGOLL

11 The heights of a group of seven-year-old boys have a mean of 121.7 cm and a standard deviation of 5.3 cm.

(Source: *www.who.int/growthref/who2007_height_for_age/en/*)

Daniel is a seven-year-old boy with a standardised score for height of 0

(*a*) Write down Daniel's height.

121·7 cm

(1)

Syed and Timur are both seven-year-old boys.

Syed's standardised score for height is 1.4
Timur's standardised score for height is −1.6

Syed is taller than Timur.

(*b*) How much taller?

group mean 121·7 SD 5·3 standardised score

Syed. 1·4

Timur −1·6

1·4 − −1·6 = 3

5·3 × 3 = 15·9

15·9 cm

(3)

Tarik takes a sample of 4 boys from the group of seven-year-old boys.
He wants to calculate the standardised score for the **sample mean** of their heights.
He uses 121.7 cm for the mean and 5.3 cm for the standard deviation.

(*c*) Discuss whether or not it is appropriate to use

(i) 121.7 cm as the mean in the calculation of the standardised score,

eg. this is appropriate since the sample mean is approximately the same as the population mean

(2)

(ii) 5.3 cm as the standard deviation in the calculation of the standardised score.

eg this is not appropriate since the spread of the sample mean is smaller than individual values of the population.

(2)

(Total for Question 11 is 8 marks)

Examiners Key Notes

Pearson
Edexcel

Foundation

Based on their performance students should:

- Read each question fully and carefully before attempting to answer it and show working out to support the final answer
- Practice:
  - Interpreting answers to statistical calculations and diagrams in context
  - Identifying the appropriateness of calculations, diagrams and approaches in different settings
  - Making comparisons, particularly those in context. Ensure that correct statistical language is used throughout when making comparisons, just stating figures is not a comparison
  - Questions where you need to 'assess' or 'discuss the validity' of a conclusion make sure that you include this in the answer
- Give a decision when the question asks for this and support it using the information from the question, such as a relevant sampling method given the sampling frame that is intended for use
- Recognise the importance of any numbers in a question as an indication that the response requires a calculation to be done or use the number to explain or justify a conclusion
- Revise:
  - How to carry out an experimental test
  - Cumulative frequency step polygons
  - Skew in distributions and be able to interpret it in context.

Higher

Based on their performance on this paper, students should:

- Use a ruler to draw graphs, show working in 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.
- Practice
  - Interpreting statistical calculations in the context of the question.
  - Analysing plans for data collection and the subsequent diagrams and calculations together with giving statistical reasons or against the approaches suggested.
  - Extended response questions, laying out answers in bullet points

- Develop their understanding of :
  - the meaning of skew.
  - conditional probability and practice a range of questions.
  - relative risk.
  - regression equations and the interpretation of the gradient of these.

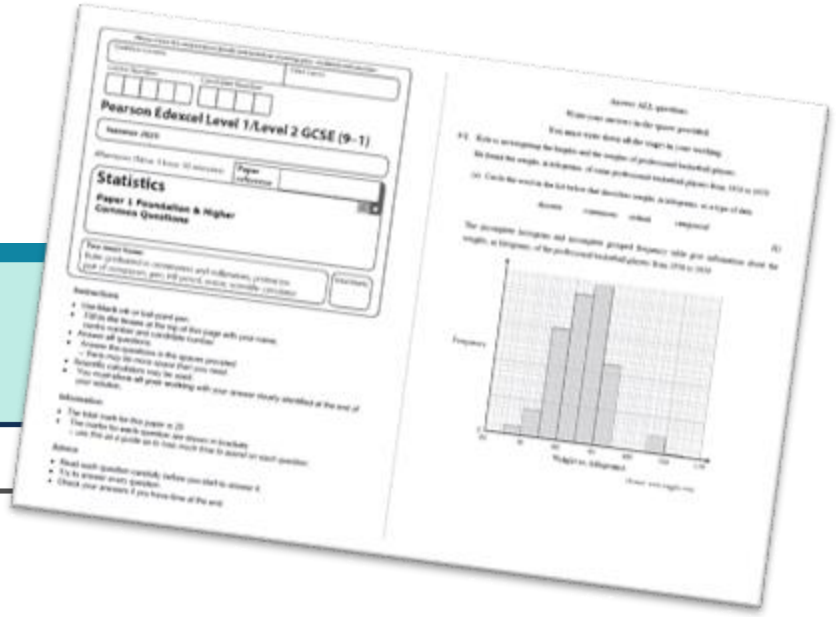
- Revise:
  - Geometric means for a range of different contexts and giving interpretations of these.
  - The names of different types of sampling and learn to identify a population
  - Standardised scores
  - Skew in distributions.

Support

Pearson
Edexcel

Maths Emporium

Do read the examiners reports.



Maths Emporium > GCSE Statistics

Category: GCSE Statistics

GCSE Statistics documents for the current 9-1 specification (1ST0)

- 01 Specification
- 02 Teacher Guidance
- 03 Schemes of Work and mappings
- 04 Specimen papers and SAMs
- 05 Past Papers & Mark Schemes
- 06 Examiner Reports and exemplars
- 07 Grade Boundaries & Skills Maps
- 08 Practice Papers & Mark schemes
- 09 Shadow Papers

mel@justmaths.co.uk

christian@justmaths.co.uk

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.