



Mark Scheme (Results)

Summer 2025

Pearson Edexcel GCSE (9 – 1)
In Statistics (1ST0)
Higher Paper 2H

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2025

Question Paper Log Number P76387A

Publications Code 1ST0_2H_2506_MS

All the material in this publication is copyright

© Pearson Education Ltd 2025

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.

- 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.

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

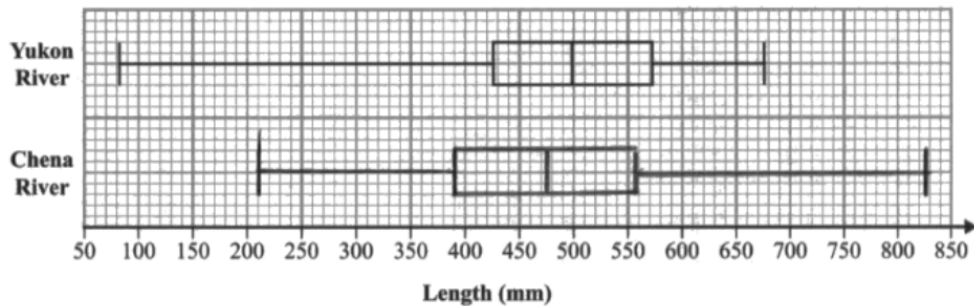
Question number	Answer	Additional guidance	Mark
1 (a)	B1 Theoretical (probability) / expected (probability)	B1 Condone misspelling	(1)
(b)	B1 B1 for any two of the following - Elijah/50% has assumed/predicted the coin is fair / Elijah assumes an even chance to land on heads/tails /Elijah is just making a prediction/ 50% would be the outcome for an unbiased coin. - Peter's/46% means the coin is biased/might not be fair - Peter's/46% was the reality/actual/experimental/calculated/worked out probability/result of experiment/test - Peter has (calculated) the experimental probability / Peter thinks heads is less likely based on the data collected/trials/flips - The experimental probability is lower than Elijah's / 50% /theoretical probability. 46% is the <u>experimental probability</u> - As the number of flips increases the probability of a heads will tend to this theoretical Elijah's 50%/probability - The observed outcomes seem to match expected frequencies(assuming the coin is fair/trial is unbiased)/the experimental probability and theoretical probability are similar/ 46% is close to 50% because we would not expect to get exactly 50% from an experiment/data/flips - Less heads were flipped (than tails) so (46% means) the coin may be biased/did not achieve the expected (150) result 46% is not close to 50% because the number of trials/flips was (too) small	B1B1 for each of 2 correct statements commenting on the difference between the theoretical and experimental probability It is possible to award B1B1 for a single comment eg '50% was based on what Elijah expected to happen, 46% came from the data' - the experimental probability is lower than the theoretical probability which means the coin/flipper could be biased '46% is the probability of heads from the experiment, 50% was for a fair coin.' '50% is for an even chance, so 46% may mean the coin is biased' Condone the use of Elijah's/50% for theoretical probability Condone the use of Peter's/46% for experimental probability unless underlined Award a maximum score of B1 per bullet point Ignore reference to one probability being more reliable/accurate than the other	(2)

Question number	Answer	Additional guidance	Mark
2	B1 for a correct comment on the hypothesis eg - the hypothesis suggested is a question not a statement (so is inappropriate). B1 B1 B1 B1 for comments on data collection Sampling method - no indication of how (Elise) will sample the skulls from the museums/we do not know if this will be primary data (that she collects/measures herself) or secondary (given the data by museum staff/records) - collecting skull measurements from museums is appropriate because it is easier/reliable/accurate/quicker OR collecting the data from all the museums may take a long time OR museums may have biased/non-representative samples - the sample size per mammal/species is small (only two males/females per species) so not representative / she should collect more female and males samples per species OR. males and females for each species often differ in size (so it is sensible to include both)/collecting data on male and female per species is representative - only collecting data on adults of the species and so she needs to include young mammals/needs a variety of ages - The species of mammals is not specified/ambiguous (Marsupials/Monotremes/Placentals etc so data may be skewed/not representative) - Large sample (100 mammals) is appropriate - Elise has made no plan to collect data on body mass (as per her hypothesis) - Elise also needs to record the type of seeds and fruits (as per her hypothesis) not just the size. Issues with the collected data - there may be errors in recording of measurements/the data may not be reliable/accurate (in the databases) OR collecting data from the existing databases/museums is likely to be a reliable/accurate source - The skulls in the museum may be damaged/degraded (so cannot get data from them) - the data may need 'cleaning' before it can be processed/data may be in the wrong format - there may be missing data in the <u>databases</u> (too)	B1 for a comment on the hypothesis B1 for each of up to four comments on data collection Award a maximum score of B1 per bullet point Contradictions within the same bullet point written scores B0 Do not accept 'Data from museum/internet may be out of date'	(5)

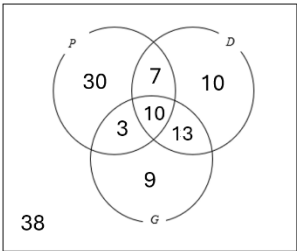
	Use of secondary data • do not know how the internet data was collected (may not be reliable/accurate/may be incorrect/biased) • the sources of the secondary data should be included / we do not know which museums/databases (she is using) OR secondary data is appropriate if a reliable source is used • skulls / seeds and fruit may not be measured by the same people or in the same way or the measurements need to be consistent and measure what was intended • Using secondary data is appropriate due to the difficulty of collecting this (“missing” primary) data (herself)/impossible to collect all information as this is primary data OR finding/collecting all the data required will be difficult/time consuming/ Elise may find it difficult to access all the data required (from the museums/databases) / databases might cost money to access OR using secondary data is appropriate because it will be cheaper/quicker/easier	Do not accept ‘Data from museum/internet may be out of date’	
--	---	---	--

Question number	Answer	Additional guidance	Mark
3 (a)	B1 Secondary (data)	B1 cao	(1)
(b)	B1 e.g. • Clippers won more games/matches than Lakers (each year/season) • Lakers won fewer games/matches than the Clippers (each year/season) • Clippers performed/ were/did better than the Lakers (each year/season) • Clippers have won the most games (each year/season) • Bars for clippers are higher/taller so did better (than Lakers each year/season)	B1 for a comparison made between the two teams for each year. Condone any use of the word ‘average/mean’ in candidates response B0 if a comparison is made of just one season e.g. 2015-2016 or comparison of each team’s performance from one season to another. Note: In this question we can ignore any figures.	(1)
(c)	B1 e.g. • (The number of games won by the Lakers year on year is) increasing/improving/rising/upward (trend)/ positive (trend) • The Lakers are winning more games each season than in the previous. • The Lakers are getting/doing better each season.	B1 for a correct comment made on the overall trend seen. eg B1 for Lakers won more and (then) more games eg B0 for Lakers won more games Do not allow positive correlation alone, but condone if accompanied by e.g. increasing /upwards trend Ignore reference to figures.	(1)
(d)	B2 for 57 and 53 (B1 for 57 or 53)	B2 for awrt 57 and awrt 53 (B1 for awrt 57 or awrt 53) Note: The reference to round to nearest whole number in the question was merely to assist candidates but is not a requirement for the B1 marks. For 57 allow 56.5(217...) For 53 allow 52.5(641...)	(2)

(e)	B1 e.g. • Clippers (because) they won more (54(%)) (of their games) whereas the Lakers only won 52/53(%). • (54(%)) > 52/53(%) Clippers won more (games). • Clippers won more games by 1%/2% • Clippers won more (44(.28)) games whereas Lakers only won 42(.64)/43 games • Clippers won 44 games which is 1 / 2 more (than Lakers) • Lakers won 42/ 43 games which is 1 / 2 less (than Clippers) • Clippers (won) 54%, Lakers (won) 52%/53% so Clippers (won more)	B1 for correctly comparing the percentage of games won by both teams with correct conclusion Note: for the Lakers, accept a value stated between 52 and 53% inclusive. A comparison of the (54(%)) and 52(%)/(53(%)) must be seen or implied as part of their answer. Condone missing 54% in their comparison if their 52/53(%) is seen and correct answer is therefore implied. 52/53(%) may be indicated on the graph Just stating figures is B0.	(1)
-----	---	---	-----

Question number	Answer	Additional guidance	Mark
4 (a)	M1 80 – ‘24’ (= 56) or sight of a value between 23- 24 A1 56 or 57	M1 for partial method. Where their ‘24’ is a value between 20 and 40. A1 56 or 57 M1A0 for a non-integer value between 56 - 57	(2)
(b)	B1 e.g. • Easy (to do)/convenient • Takes less time/more efficient • Can change location/flexible (along the river)	B1 for a correct comment about the advantages of using opportunity sampling B0 for cheaper or more accurate.	(1)
(c)	M1 for a box with two whiskers AND at least two values plotted correctly AND an attempt at the median line.  A1 for all correct	M1 for box with two whiskers and an attempt at a single median line AND at least two values plotted correctly A1 Fully correct plot (210, 390, 476, 556, 825)	(2)

(d)	B1ft B1ft B1ft B1ft		Allow ft for their box plot. Also allow recovery if candidates refer back to values in the table. For contextual comments we are looking for reference to lengths/size (of pike) Do not accept eg Yukon river is longer B1ft For a correct statistical statement comparing medians. Condone longer for comparison of medians. Must refer to <u>median</u> , not average for this B mark. Condone misspelling but medium is B0. B1ft For a correct comparison of the IQRs/ranges. Condone wider for comparison of IQR/range. B1ft For correct comparison of the skews. For Chena/Yukon also accept no skew but not neutral/normal/symmetrical skew. Condone use of ‘normally distributed’ instead of symmetric Accept ‘symmetrical distribution’ B1ft One correct interpretation Do not award mark for interpretation if clearly linked to incorrect statistical measure Allow equivalent/converse statement but underlined words must be seen. Note: in this question ignore any numerical values in comparisons	(4)								
	<table><tr><th>Comparison</th><th>Interpretation</th></tr><tr><td>Yukon (pike) greatest <u>median</u> or Chena (pike) <u>median</u> < Yukon (pike) <u>median</u></td><td>eg Yukon (pike) are larger/wider/longer/bigger (in length).</td></tr><tr><td>Chena (pike) has the greatest <u>IQR</u> or Yukon (pike) (≈ 145) <u>IQR</u> < Chena (pike) (≈ 165) <u>IQR</u> or Chena has the greatest <u>range</u> or Yukon (≈ 595) <u>range</u> < Chena <u>range</u> (≈ 615)</td><td>eg Yukon (pike) lengths are the most consistent</td></tr><tr><td>Yukon (pike) have <u>negative skew</u> /<u>are symmetrical</u> and Chena (pike) are <u>symmetrical/ have positive/negative skew</u></td><td>eg The spread of <u>lengths</u> below the average (median) for Yukon is smaller than the spread of <u>lengths</u> above the average or The spread of <u>lengths</u> above the average (median) is the same as the spread of <u>lengths</u> below the average for (both) Yukon (and) Chena. or eg the spread of lengths below the average for Chena is larger than the spread of lengths above the average.</td></tr></table>				Comparison	Interpretation	Yukon (pike) greatest <u>median</u> or Chena (pike) <u>median</u> < Yukon (pike) <u>median</u>	eg Yukon (pike) are larger/wider/longer/bigger (in length).	Chena (pike) has the greatest <u>IQR</u> or Yukon (pike) (≈ 145) <u>IQR</u> < Chena (pike) (≈ 165) <u>IQR</u> or Chena has the greatest <u>range</u> or Yukon (≈ 595) <u>range</u> < Chena <u>range</u> (≈ 615)	eg Yukon (pike) lengths are the most consistent	Yukon (pike) have <u>negative skew</u> / <u>are symmetrical</u> and Chena (pike) are <u>symmetrical/ have positive/negative skew</u>	eg The spread of <u>lengths</u> below the average (median) for Yukon is smaller than the spread of <u>lengths</u> above the average or The spread of <u>lengths</u> above the average (median) is the same as the spread of <u>lengths</u> below the average for (both) Yukon (and) Chena. or eg the spread of lengths below the average for Chena is larger than the spread of lengths above the average.
	Comparison	Interpretation										
	Yukon (pike) greatest <u>median</u> or Chena (pike) <u>median</u> < Yukon (pike) <u>median</u>	eg Yukon (pike) are larger/wider/longer/bigger (in length).										
	Chena (pike) has the greatest <u>IQR</u> or Yukon (pike) (≈ 145) <u>IQR</u> < Chena (pike) (≈ 165) <u>IQR</u> or Chena has the greatest <u>range</u> or Yukon (≈ 595) <u>range</u> < Chena <u>range</u> (≈ 615)	eg Yukon (pike) lengths are the most consistent										
Yukon (pike) have <u>negative skew</u> / <u>are symmetrical</u> and Chena (pike) are <u>symmetrical/ have positive/negative skew</u>	eg The spread of <u>lengths</u> below the average (median) for Yukon is smaller than the spread of <u>lengths</u> above the average or The spread of <u>lengths</u> above the average (median) is the same as the spread of <u>lengths</u> below the average for (both) Yukon (and) Chena. or eg the spread of lengths below the average for Chena is larger than the spread of lengths above the average.											

Question number	Answer	Additional guidance	Mark
5(a)	 B1 for 7 and 13 in correct place B1 for 30, 10 and 9 in correct place B1 for 38 in correct place	SCB2 If all are written as probabilities in the correct regions as correct fractions oe eg decimals or a mix of both but must all be probabilities. SCB1 for if correct probabilities written in correct regions for 7, 13, 30, 10 and 9	(3)
(b)	B1 $\frac{35}{120}$ oe	B1 for correct probability DO NOT allow for follow through from their Venn diagram allow equivalent fraction, eg $\frac{7}{24}$ awrt 0.29 or awrt 29% ISW if correct answer seen Note: 35:120 or 7:24 or 35 out of 120 or 7 out of 24 is B0	(1)
(c)	M1ft $\frac{10 + "13"}{"10" + "13" + "7" + 10} \left(= \frac{"23"}{"40"} \right)$ or $\frac{"10" + "13"}{120} \div \frac{"10" + "13" + "7" + 10}{120} \left(= \frac{"23"}{120} \div \frac{"40"}{120} \right)$ or "0.19(16...)" ÷ "0.33(33...)" A1ft $\frac{"23"}{"40"}$	M1ft Allow for follow through from their Venn diagram or for their overall denominator of "10" + "13" + "7" + 10 (= "40") or correct probability division seen with denominators of 120 and their "23" < "40" A1ft allow equivalent fractions, or a decimal of their "0.575" / "0.58" / "0.57" or their percentage "57.5%" / "58%" / "57%" Note: if answer is written as a percentage then percentage symbol must be given. ft their Venn diagram with probability to 2dp (rounded or truncated).	(2)

(d)	B2 ft No, (Toga is incorrect) since $P(G D) \neq P(G)$ or No, (Toga is incorrect) since $\frac{35}{120} \neq \frac{23}{40}$ OR No, (Toga is incorrect) since $P(G \cap D) \neq P(G) \times P(D)$ AND $\frac{23}{120} \neq \frac{35}{120} \times \frac{40}{120} \left(= \frac{1400}{14400} = \frac{7}{72} \right)$ (B1 ft $P(G D) \neq P(G)$ or $\frac{35}{120} \neq \frac{23}{40}$ OR $P(G \cap D) \neq P(G) \times P(D)$ AND $\frac{23}{120} \neq \frac{35}{120} \times \frac{40}{120} \left(= \frac{1400}{14400} = \frac{7}{72} \right)$)	B2 ft for a correct conclusion with correct mathematical deduction that $P(G D) \neq P(G)$ when comparing their parts answers to (b) and (c) OR $P(G \cap D) \neq P(G) \times P(D)$ with the correct fractions $\frac{35}{120}$ and $\frac{40}{120}$ used and compared to $\frac{23}{120}$ or their $\frac{23}{120}$ from the Venn diagram in (a). (B1ft for correct mathematical comparison with no or incorrect conclusion)	(2)
-----	---	--	-----

Question number	Answer	Additional guidance	Mark
6 (a)	B1 eg (Marcus might be wrong because) (The pie charts) do not show the total/population/frequency/number of livestock/actual values/exact number (The pie charts) only show the proportion/percentage (was greater for turkeys in Finland)	B1 Allow an implication that the totals for Finland and Norway may be different. Accept eg the total populations might be different / There might be more livestock in Norway Accept eg conclusions cannot be made because pie charts only show percentages/proportions	(1)
(b)	M1 $\frac{2341780}{100} \times 35$ A1 819 623 SCB1 for an answer of 468356	M1 for any full correct equivalent complete method. A1 819623 cao Correct answer only with no working scores M1A1	(2)

(c)	M1 $\frac{14\,888\,487}{2\,341\,780} = \frac{r^2}{25}$ or $25 \times \frac{14\,888\,487}{2\,341\,780} (=158.944\dots)$ or $\frac{2\,341\,780}{14\,888\,487} = \frac{25}{r^2}$ or $25 \div \frac{2\,341\,780}{14\,888\,487} (=158.944\dots)$ or $\frac{5^2\pi}{r^2\pi} = \frac{2\,341\,780}{14\,888\,487}$ or $\frac{14\,888\,487 \times 25\pi}{2341780\pi} = r^2$ or better and equivalent A1 12.607(30446...) (cm)	M1 for a correct equation connecting the radii - squared or for a complete method to find the radius. or $14\,888\,487 \div \frac{2\,341\,780}{25}$ or $\frac{14\,888\,487}{r^2} = \frac{2\,341\,780}{25}$ or $14\,888\,487 \times 25 = 2\,341\,780 \times r^2$ or better and equivalent $\frac{r}{5} = \frac{\sqrt{14\,888\,487}}{\sqrt{2\,341\,780}}$ or $5 \times \sqrt{\frac{14\,888\,487}{2\,341\,780}}$ or $5 \div \sqrt{\frac{2\,341\,780}{14\,888\,487}}$ oe or $5 \times 2.521460\dots$ or $5 \div 0.396595\dots$ A1 awrt 12.6 M0A0 for 31.8 (or 32)	(2)
-----	--	--	-----

Question number	Answer	Additional guidance	Mark
7(a)	M1 $0.9 + 0.6 + 0.4 + 0.2 + 0.1$ A1 $= 2.2$ or $\frac{11}{5}$	M1 for adding the five appropriate percentages together. A1 2.2 oe ISW attempt to simplify / convert if correct answer seen.	(2)
(b)	Method: Proportional amount added to LB using exclusive method M1 $30 + \left((50 - 32.04) \times \frac{45 - 30}{19.52} \right) (= m)$ or $30 + \left((45 - 30) \times \frac{50 - 32.04}{19.52} \right) (= m)$ A1 43.801(22951..) OR using inclusive method: M1 $30 + \left[((101 \div 2) - 32.04) \div 19.52 \right] \times 15 (= m)$ oe A1 44.18545(082...) OR Method: Proportional amount subtracted from the UB using exclusive method M1 $45 - [(51.56 - 50) \div 19.52] \times 15 (= m)$ A1 43.801(22951..) (OR using inclusive method: M1 $45 - [(51.56 - 50.5) \div 19.52] \times 15 (= m)$ A1 44.18545(082...)	M1 for correct method of interpolation using the tabulated data using the exclusive/inclusive method, working from either LB up, or UB down. allow alternative form/simplification eg $\frac{50 - 32.04}{19.52} = \frac{m - 30}{45 - 30}$ or $\frac{50.5 - 32.04}{19.52} = \frac{m - 30}{45 - 30}$ eg $30 + ((50 - 32.04) \div 19.52) \times 15 (= m)$ or $30 + ((50.5 - 32.04) \div 19.52) \times 15 (= m)$ or $\frac{17.96}{19.52} = \frac{m - 30}{15}$ or $\frac{18.46}{19.52} = \frac{m - 30}{15}$ OR eg $\frac{51.56 - 50}{19.52} = \frac{45 - m}{45 - 30}$ or $\frac{51.56 - 50.5}{19.52} = \frac{45 - m}{45 - 30}$ A1 (dep M1) awrt 44	(2)

(c)	B1ft eg • The median falls in the interval $30 \leq y < 45$ / 30 to 45 for 2023, so yes – Kiran’s claim is correct. • In 2023, 28.6% of population was aged 27 or younger. 28.6% < 50% , so yes – Kiran’s claim is correct. • In 2023, 71.4% of population was aged 27 or older. 71.4% > 50% , so yes – Kiran’s claim is correct. • Yes, because the median has increased by “17” years • 44 (years)...is greater than 27 (years) so Kiran’s claim is correct • 43.8 > 27 so Kiran’s claim is correct	B1ft for a correct comparison with a conclusion. Allow a comparison with the tabulated data and comparing to an interval the median falls within. Be aware, 44 may be written as 43.801(22951..) or 44.18545(082...) from part (b) which would have been correct alternatives to 44 Award B1ft here. B0 for “correct as median is higher in 2023 than 1960” or “no, the median is decreasing showing less older people”. A clear comparison must be made.	(1)
(d)	B1eg • Loss of accuracy (by working with grouped data / interpolating/raw data lost/actual data lost) • (linear interpolation from the) table class intervals has assumed the data is evenly spread • The website might have used different class intervals/no intervals/raw data/actual data	B1 Comment explaining the implication of merging or assuming evenly spread data B0 for “linear interpolation is an estimate” only	(1)

Question number	Answer	Additional guidance	Mark
8 (a)	B1 for eg - data is bivariate / paired / ‘because <u>each athlete/person has 2</u> pieces of data recorded’ - Allows Zara to see if there is a correlation / relationship / association - So Zara can see if height influences/affects the long distance jumped	B1 for justifying the appropriateness of a scatter diagram. Underlined words are needed. Allow miss-spellings if underlined word is clearly attempted. Ignore positive/ no/ negative label to correlation for second bullet point. Accept <u>bivariable</u>	(1)
(b)	B1 e.g. - height is the independent/explanatory/predictor variable (so should be plotted on the x-axis) - height is not the response/dependent/predicted variable - height influences the long jump distance - long jump distance relies on height/the jump is the result of how tall they are - The (height) may affect the distance jumped	B1 for knowing and applying the term explanatory (independent) variables.	(1)
(c)	B1 - for no (linear) correlation. - for almost no correlation .	B1 Know that the PMCC measures the strength (of a linear) correlation. Do not accept neutral. Condone (weak) positive correlation	(1)

(d)	B1 for negative (linear correlation) B1 for as the long <u>jump</u> (distance) increases/improves the (100m) <u>race/running</u> time decreases/(race time) improves eg “as the long jump distance increases the race time decreases.” or “as the race time increases the long jump distance decreases.” or “As the long jump distance improves, the race time improves” or “If the race time is slower the long jump distance is shorter.	B1 for a correct interpretation of the correlation Ignore any references to strength B1 for a correct contextual interpretation Note: The correlation is negative, but the context is also inversed. Longer distances in long jump imply better results. Shorter times in running implied improved results. Note: Accept converse when describing the correlation eg “as the <u>race time</u> decreases the <u>long jump</u> (distance) increases” Underlined words are needed.	(2)
(e)	B1 for interpolating / inside range of race times dep B1 ...so reliable OR B1 for sample size is (very) small depB1 ...so not reliable	B1 for knowing and applying vocabulary of interpolation dep B1 concluding that this is reliable. OR B1 for assessing the appropriateness of statistical methodology dep B1 concluding that this is unreliable. Maximum B1 for a candidate who has concluded it is both unreliable and reliable for alternative reasons	(2)

Question number	Answer	Additional guidance	Mark
9(a)	B1 eg - The results will be <u>reliable</u> because the sampling method avoids any bias / these control measures will result in <u>reliable</u> data for comparison/data analysis - The results <u>are reliable</u> since the samples were collected at the <u>same time</u>/were all sent to the <u>same laboratory</u>/(collected for) <u>10 days/same number of days</u> / (100ml) <u>same amount</u> collected each day OR - The results <u>are unreliable</u> since samples only taken at one time in the day / 10 days is a small sample size / samples should be tested at more than one laboratory	B1 for a comment on the reliability of their results with a logical reason. OR B1 for a comment on the unreliability of their results with a logical reason. B0 for a response that concludes the control variables are reliable but also unreliable. B0 for vague comment on extraneous variables	(1)
(b)	M1 $7 + (3 \times \sigma) = 9.4$ A1 $(\sigma =) 0.8$ oe OR M1 (upper action limit – upper warning limit =) 9.6 – 8.8 A1 $(\sigma =) 0.8$ oe	M1 For correct calculation seen for the upper action limit to find the standard deviation A1 (dep M1) for correct or equivalent answer OR M1 For using the upper action limit and upper warning limit from the graph below this item and finding the difference between them. $(\mu + 3\sigma) - (\mu + 2\sigma) = \sigma$ A1 for correct or equivalent answer M0A0 for 0.8 written on the answer with incorrect working	(2)

(c)	B1 e.g. • 10 > 9.4 (which is 3 standard deviations above the mean) so it is very unlikely (for this level of lead) to occur. • 10 falls above 99.7% /99.8%, so unlikely • The probability of it being 10 is less than 0.15%/0.1% so unlikely • 10 is a lot higher than the upper action limit so unlikely • It is unlikely because 10 would have stopped the process immediately • a point beyond the action limit would have meant the laboratory would have stopped the process and checked the sample/testing equipment. • 10 would mean there is a fault with the sample/testing equipment	B1 for knowledge that for a normal distribution, values more than three standard deviations from the mean are very unlikely/unusual Condone 0.3% / 0.2% for 3rd bullet point. Condone the use of “almost all” the data rather than writing the 99.7%/99.8% eg “almost all the data should fall within 3 sd of the mean so unlikely” Condone use of “impossible” instead of unlikely. OR for commenting on the laboratory taking action	(1)
(d)	B2 for eg • All/the sample_means and (all/the) sample_ranges are within <u>warning</u> limits (B1 for e.g. • Sample means are within <u>warning</u> limits • Sample ranges are within <u>warning</u> limits)) B1dep No action is required [dependent on B2]	B2 for referring to both the sample means AND the sample ranges being within <u>warning</u> limits. (B1 for referring to the sample means OR the sample ranges being within <u>warning</u> limits) B1dep dependent on B2 For a correct conclusion Note: For B2 and B1 accept “<u>sample means</u> and <u>sample ranges</u> do not exceed any limits” Do not accept “does not exceed action limits” alone. Condone “ nothing/none of the data exceeds any limits” B2depB1	(3)

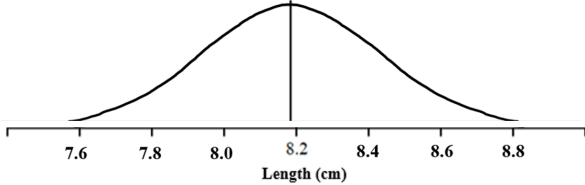
Question number	Answer	Additional guidance	Mark
10(a)	M1 $222000 / 224000 \times 100$ A1 99.1 M1 for $\sqrt[5]{111.1 \times 106.7 \times 105.5 \times 105.4 \times "99.1"}$ = 105.4894531 A1 ft 105.5 depB1ft (The average) increase is “5.5%” <u>per year</u>	M1 for correct calculation of chain base index number A1 for awrt 99.1 Note: 99.1 seen implies M1A1. 99.1 may be written in the table. Condone £99.1(0) M1 for correct calculation of the geometric mean of 5 chain base index numbers SCB1 for a candidate who calculates the geometric mean of the 4 chain base numbers in the table. $\sqrt[4]{111.1 \times 106.7 \times 105.5 \times 105.4}$ (=107.150(1756 Candidate may go on to earn A0B1ft A1 ft for awrt 105.5 Follow through their 99.1 The arithmetic mean is 105.56 Please ensure the method is fully correct before awarding the second A mark. depB1ft dependent on M1 for correct calculation of geometric mean of 5 (or 4) chain base index numbers. For a correct interpretation of the value of the geometric mean	(5)

(b)	M1 $3 \times \frac{235884-213250}{119492}$ A1 0.5682...= 0.6 (So), positive (skew)	M1 correct use of Pearsons Skew formula. A1 awrt 0.6 AND stating (the skew is) positive or a correct interpretation of positive skew is given. Allow correct interpretations in context. eg ‘mean > median’ or ‘there are more extreme high values (pulling the tail) out to the right. or ‘most of the data lie below the mean, but a few large values drag the average upward’ Do not accept alone... “more spread above the mean” or “more data above the median” to imply positive skew For a response that calculates 0.6 and states positive (skew) but then goes onto give an incorrect contextual interpretation score M1A1 We will ignore an incorrect context given if 0.6 and positive stated.	(2)
-----	---	--	-----

Question number	Answer	Additional guidance	Mark
11(a)	M2 for $\frac{(32.5 \times 25) + (37.5 \times 18) + (42.5p) + (47.5 \times 14) + (52.5 \times 3)}{25 + 18 + p + 14 + 3} (= 40.5)$ (M1 for $(32.5 \times 25) + (37.5 \times 18) + (42.5p) + (47.5 \times 14) + (52.5 \times 3) (= 2310 + 42.5p)$ A1 $p = 60$	M2 for a correct expression for the calculation of the mean, using the midpoints and an implied attempt to sum the components of the numerator and the denominator. May be seen as $\frac{812.5 + 675 + 42.5p + 665 + 157.5}{25 + 18 + p + 14 + 3} \text{ or } \frac{2310 + 42.5p}{60 + p}$ M1 for a correct expression for the sum of the numerator OR the denominator. Allow M1 for consistent use of fx with x within interval (including end points) and sum. Or allow for a maximum M1 one error in their calculation if they have used midpoint values. The following oe all imply M2 $2310 + 42.5p = 40.5(60 + p)$ $2310 + 42.5p = 2430 + 40.5p$ $42.5p - 40.5p = 2430 - 2310$ $2p = 120, p = 120 \div 2$ A1 for $p = 60$ $p = 60$ given with no working award M1M1A1	(3)
(b)	B1 e.g. - we do not know the total number of cars that have been surveyed, only those that exceeded the speed limit. - We cannot work out the percentage of cars that broke the speed limit	B1 For a comment that clearly states the limitations of analysis of the data in the table.	(1)

Question number	Answer	Additional guidance	Mark
12(a)	B1 for 22.85 – 21.33 (or sight of 1.52) Girls B1 for 23.45 – 21.86 (or sight of 1.59) Boys depB1 for eg - girls IPR is smaller (than boys IPR) $1.52 < 1.59$ so girls (IPR) is smaller For the award of this mark the boys IPR and girls IPR must be clearly identifiable. ALT B1 5th percentile B – G 21.86 – 21.33 (or sight of 0.53) B1 75th percentile B – G 23.45 – 22.85 (or sight of 0.6) depB1 for eg - Hence Boys IPR > Girls IPR $0.6 > 0.53$ so boys (IPR) is larger	B1 for correct calculation of girls IPR B1 for correct calculation of boys IPR depB1 for comparison and correct conclusion. depB1 dependent on B1B1 For depB1 allow “The eye length of the boys has increased by 0.07 more than the girls” Allow converse statements for depB1 ALT B1 for correct calculation of difference between the girls 5th and the boy 5th percentile. B1 for correct calculation of difference between the girls 75th and the boy 75th percentile. Note: the candidate might do G – B instead which is fine, but both need to be the same direction of subtraction depB1 for comparison and correct conclusion. depB1 dependent on B1B1 Allow converse statements for depB1 Condone range instead of IPR Condone wider spread for greater IPR	(3)

(b)	B1ft for eg • (the eye length of 6 year old) girls are more consistent/less spread out/less varied (than the eye lengths of (6 year old boys). • (the eye length of 6 year old) boys are less consistent/more spread out/more varied (than (6 year old) girls eye lengths).	B1ft for a correct interpretation about the spread of the data Follow through their IPR values from (a) Allow “greater range” or the use of the word range to imply a greater breadth or broadness.	(1)
------------	---	---	------------

Question number	Answer	Additional guidance	Mark
13 (a)	B1 $(8.6 - 8.2) \div 2 (= 0.2)$ Or $8.2 + (0.2 \times 2) = 8.6$	B1 for a correct calculation seen	(1)
(b)	B2 bell shaped curve centred on 8.2 and tails ending around 7.6 or a lower value for the LH-tail and around 8.8 or a higher value for the RH-tail  (B1 bell shaped curve centred on 8.2)	B2 There must be an attempt at a symmetrical bell-shaped curve centred around 8.2 and at least one value indicated on the axis to emphasise the tails positioned at 7.6 or a lower value (no lower than 7.4) and 8.8 or a higher value (no higher than 9). Condone the tails being labelled as 7.6 and 8.8 at different points on the axis. i.e. they can use they can scale the axis however they deem fit. (B1 for an attempt at a symmetrical bell-shaped curve centred around 8.2 but with no or incorrect scale applied to axis to indicate correct position of tails or for a non-symmetrical bell shaped curve (but a clear attempt at a bell shaped curve) that ends at 7.6 and 8.8) Condone the curve just touching the x-axis at their tails or touching the x-axis at 7.6 or 8.8 for B2. Maximum award of B1 for a curve that crosses and goes below the x – axis.	(2)
(c)	B1 e.g. • ‘unlikely’ (since) 7.5 cm is an outlier • ‘unlikely’ (since) 7.5 cm is smaller than 7.6 • ‘unlikely’ (since) 7.5 cm is smaller than $\mu - 3\sigma$/over 3 sd (away) from the mean • unlikely as only 0.1% of data below 7.6	B1 for indicating that the 7.5 cm is an unlikely (or equivalent phrase) and stating this is because 7.5 cm is an outlier or some other equivalent reason.	(1)

Question number	Answer	Additional guidance	Mark
13 (d)	<u>Tabulation</u> Comments to support Liam's tabulation being appropriate: B1 B1 for any two from (Appropriate since...) • they do not overlap • equal class intervals are suitable if data is closely grouped within a range of values and not more spread out in others • as normally distributed data would be more closely clustered in the middle of the range of values • The data collected has considered all possible ages/the data is exhaustive (since no butterfly lives longer than 15 days) OR Comments to support Liam's tabulation being inappropriate: B1 B1 for any one from (Inappropriate since...) • (for a normal distribution) Less data is collected for the extreme class intervals, so the groups would be merged for lower and higher group intervals. • (for normal distributions) the class intervals do not need to be the same width/ frequency densities would be more appropriate for different sized class intervals. • more class intervals should have been included • more intervals in the middle/centre of the table/data/distribution would make enable him to assess/analyse skew	B1 B1 for two comments on the class intervals. Candidate can give two reasons for the tabulation being appropriate Or Two reasons for the tabulation being inappropriate Or One reason for it being appropriate and one reason for it being inappropriate – provided this is not a direct contradiction.	(5)

	<u>Histogram</u> B1 for any one from (Appropriate since...) • Histogram will allow you to see if the distribution is bell-shaped • Histogram will allow Liam to see if the data is skewed • Histogram requires continuous data • Histogram appropriate as data is grouped Or B1 for (Inappropriate since...) • sample size is too small to be representative. <u>Calculations</u> B1 for any one from (Appropriate since...) • the mean and median should be (approximately) the same value (for a normal distribution) • if the mean and median are the same the distribution is (likely to be) symmetrical / if the mean and median are not the same the distribution is (likely to be) skewed (so not normally distributed) • Comparing mean and median will allow him to see if data is skewed B1 for any one from (Appropriate since...) • 68% of the data should lie within 1 standard deviation of the mean • 95% of the data should lie within 2 standard deviations of the mean • 99.7/8% of the data should lie within 3 standard deviations of the mean	B1 for comment on the use of a histogram Note – candidate may give a reason for the histogram being appropriate or inappropriate Do not award if response goes on to contradict itself and therefore seems undecided over appropriateness. B1 for comment on the mean and median expected to be about the same size if the distribution is normal B1 for comment on percentage of data expected within 1 or 2 standard deviations of the mean. Note – candidate is expected to agree with the appropriateness of the calculations to check for characteristics of a normal distribution, but they can give reasons for the tabulation and histogram being inappropriate.	
--	--	--	--

