

Mark Scheme (Results)

Summer 2022

Pearson Edexcel GCSE In Statistics (1ST0) Higher Tier 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 2022
Question Paper Log Number P68676A
Publications Code 1ST0_2H_2206_MS
All the material in this publication is copyright
© Pearson Education Ltd 2022

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.
- 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	Answer	Additional guidance	Mark
1 (a)	M1 0.452 × 31 (= 14.012) or $\frac{d}{31}$ = 0.452 A1 14 (days)	M1 for correct use of absolute risk implied by awrt 14.0 A1 for cao (must be an integer) Answer only is 2 out of 2	(2)
(b)	M1 0.452 × 1.29 × 31 (= 18.075) A1 18 (days)	M1 for correct use of relative risk Allow their '14' × 1.29 (=18.06) May see 0.583 × 31 for M1 implied by awrt 18.1 A1 for cao (must be an integer) Answer only is 2 out of 2	(2)
		Note: in (a) and (b) penalise lack of rounding to integer once only, i.e. if answers are awrt 14.0 and awrt 18.1 then score M1A0M1A1. SC: Answer of 15 in part (a) and 19 in part (b) scores M1A0M1A1	

Question	Answer		Additional guidance	Mark
2 (a)	M1 A1 A1 150 155 160 165	170 175 180 185 190 195	M1 for box with at least one whisker with at least one correct value 1st A1 for 3 correct values plotted from 156, 166, 170, 176, 182	(3)
(b)	B1 B1 B1 B1		B1 Correct comparison of medians B1 Correct comparison of spread (IQR or	(4)
	Comparison Basketball median > rugby median Basketball IQR = rugby IQR Basketball range < rugby range Basketball no skew/ symmetric(al) and Rugby positive skew	Interpretation Basketballers are taller (on average) Similar variation in height Basketballers' heights (slightly) more consistent Rugby players more varied above median (and below upper quartile)/basketball same spread either side	range) B1 One further correct comparison B1 One correct interpretation Allow equivalent/converse statements but underlined words must be seen. Do not allow 'normally distributed' or 'symmetrical skew' for symmetrical For skew comment also allow alternative comparison considering whiskers. (i.e. both show negative skew)	
(c)	B2 Comparisons may be unreliable because (any one from) • No indication that samples were random/representative • Based on secondary data/reliability of sources unknown • Samples were small /only based on 15 players		B2 for unreliable with a correct supporting reason (B1 for unreliable with attempt at reason OR appropriate consideration of reliability without a conclusion) SC: For B2 allow 'reliable since data comes from trustworthy website' (oe)	(2)

Question	Answer	Additional guidance	Mark
3 (a)	B1 e.g. countries with higher percentage of urban population have higher life expectancy	B1 for an appropriate statement linking urban population (oe) and life expectancy (oe). Condone e.g. 'people living in urban areas live longer' for this mark. A question scores B0.	(1)
(b)	B1 Either: Data is paired / bivariate Or: Scatter diagram will show any correlation	B1 for an appropriate justification for a scatter diagram. Underlined words are needed.	(1)
(c)	B1 Urban population (%) is the explanatory variable depB1because Irina believes this affects life expectancy, orbecause life expectancy is determined by this, orbecause life expectancy is the response variable	B1 for identifying the explanatory variable. B1 (dependent on previous B1) for correct reasoning. Accept equivalent comments if meaning is clear	(2)
(d)	B2ft e.g. positive correlation, so hypothesis is supported	B2ft for correct conclusion with reference to positive correlation and consistent with their hypothesis in (a). If B0 scored in part (a), then max score is B1 (B1ft for an incomplete answer, either missing vocab. OR correct reasoning with incorrect/no conclusion)	(2)

(e)	B2 Line of best fit drawn through (63.9, 77.8)	B2 for appropriate line of best fit with positive gradient through given mean point. (Mean point need not be plotted).	(2)
		(B1 for plotting the double mean point or for an appropriate line with positive gradient not through double mean point)	
(f)	B1 e.g. for every extra 1% in urban population , life expectancy increases by 0.19 years (accept appropriate use of their gradient if found)	B1 for a complete equivalent numerical interpretation of gradient including population (oe) and years (oe)	(1)
(g)	 B1 For any one from lower life expectancy value than expected (65, 63) plotted correctly life expectancy for 65% should be around 77/78 (years) life expectancy read off their line of best fit at 65% 	B1 for correct reasoning	(2)
	B1 Anomaly / does not fit with other data	B1 for correct conclusion Condone 'outlier'	

Question	Answ	ver	Additional guidance	Mark
4 (a)	B1	Random numbers are allocated in the same proportion (oe) as (average) number of consultations.	B1 for indicating the relationship between random numbers and number of consultations. Implied by one numerical comparison e.g. $\frac{32}{100} = \frac{16}{50} \text{OR } 16:24:10 = 32:48:20$ OR random numbers are double (average number of) consultations (per consultant), etc Once B1 has been scored, ignore any incorrect numerical comparisons seen.	(1)
(b)	B1	A = 21, B = 21, C = 8	B1 cao	(1)
(c)	B1ft	$[(21' - 8') \times 20 =]$ 260 minutes	B1 for 260 or ft their 21 and 8 from (b) for consultants A and C	(1)
(d)	B1 B1	Simulation must be repeated / each simulation will give a different result Any one from: • we need to see the distribution of results • to identify likely maximum/minimum • to see how much the results vary • assumes there are always 50 consultations (every week) • a comment relating to the time of each consultation. e.g. 'the simulation doesn't measure the time of each consultation (just the number of consultations)'	B1 for recognising that repetition is needed / results will differ if repeated. B1 allow equivalent wording which recognises the need to see the variation in results. Accept equivalent to any one bullet.	(2)

Question	Answer	Additional guidance	Mark
5 (a)	 B1 Any one from: Allows for comparisons (between control group and test group) Controls for extraneous variables 	B1 for an appropriate equivalent reason for using a control group implying that comparisons can be made.	(1)
(b)	B1 Will aim to pair people (placing one in each group) •with similar fitness / initial energy levels, or •of similar age / same gender / similar characteristics	B1 for an appropriate equivalent statement of how matched pairs should be selected (Do not allow reasons why matched pairs should be selected)	(1)
(c)	EITHER: B1 Conclusions may not be reliable, due to B1 B1 any two sensible reasons, e.g.: • sample size of 10 is small • relies on subjective judgement (of level of fitness / energy levels) or accuracy of questionnaire • the two groups may not stick to their own drink / use different amounts of drink • may not be possible to have two equivalent groups • may not all have same physical engagement on expedition (e.g. carrying different weights) • the placebo/psychological effect OR: B1 Conclusions may be reliable, as B1 B1 any two sensible reasons, e.g.: • there is an objective test of fitness/energy levels • the two groups stick to their own drink / use same amount • control groups/matched pairs are used • all have same physical engagement on expedition (e.g. carrying same weights)	1st B1 for a conclusion about reliability (either supporting or refuting) with an attempt at justification. 2nd B1/3rd B1 for up to two comments which can be used to justify a comment about reliability. Do not accept comments inconsistent with reliability statement. If both reliable and unreliable are assessed, then score the highest.	(3)

Question	Answer	Additional guidance	Mark
6 (a)	B1 For any one from:	B1 for an indication of size or randomness of the sample	(1)
	 (Each) sample needs to be random/each trout has the same chance of being caught Trout should be allowed to mix before taking second sample 	Ignore extraneous comments	
	• (Each) sample should be large enough (to be representative)	Comment must be referring to samples (not populations)	
(b)	M1 $50 \div \frac{16}{80}$ or $\frac{50}{n} = \frac{16}{80}$ o.e. A1 250 (trout in the lake)	M1 for a method that would lead to a correct estimate. A1 cao	(5)
	Statement Condition/assumption Only valid iftags remain in placepopulation is 'closed' (no births/deaths etc)sampling method is same each timetagging fish does not affect the likelihood of being recapturedappropriate time between the two samplesrandom sampling is used Reliabilityis high as sample sizes are large	B3 for a statement about validity or reliability with two conditions/assumptions from those listed (B2 for a statement about validity or reliability with one condition/assumption from those listed OR reference to any two conditions/assumptions without mention of validity or reliability) (B1 for reference to any one conditions/assumption from those listed with no reference to validity or reliability) Allow converse statements. Note, for statements allow use of 'validity' for 'reliability' and vice versa, but do not allow 'accurate'.	

Question	Answer	Additional guidance	Mark
7 (a)	 B1 Any one from Quicker to collect lots of data compared with primary Less expensive (than collecting it yourself) Convenient/easier to collect/access the data 	B1 for any advantage of using secondary data Allow a disadvantage of using primary data as long as it is clear that they referring to primary data	(1)
(b)	M1 $16 \times 5 (=80)$ or $13.6 \times 5 (=68)$ M1 $'80' + '68' + 14$ A1 $= 162$	M1 for correct use of area as frequency (may be implied by one correct area and may be seen on histogram) M1 for complete method	(3)
		A1 cao Condone: 1.4 × 10 (=14) 4.8 × 10 (=48) 0.3 × 50 (=15) 238 – ('14' + '48' + '15') = 161 for M1M1A1. An answer of 161 or 162 scores 3 out of 3. SC: An answer of 147 or 148 scores M1M1A0.	
(c)	M1 e.g. $\frac{238 - 162}{238}$ (= $\frac{76 \text{ or } 77}{238}$) or $\frac{2}{3} \times 238$ or $(238 - 162) : 238$	M1 for any correct (or ft) calculation to allow comparison with $\frac{1}{3}$ or with '162'. Condone: $1.4 \times 10 + 4.8 \times 10 + 0.3 \times 50$ (=77)	(4)
	A1ft = awrt 0.32 or $76/77 < 79(.3)$ or =awrt 159 or = 1 : awrt 3.1	Allow equivalent alternatives Alft for any correct answer (ft their 162)	
	A1ft so Andy is incorrect	A1 ft for correct conclusion (ft their 162) Condone use of 239 instead of 238	
	B1 e.g. May not be reliable as this week may not be typical (or unknown reliability of secondary data, or excludes weekends), OR e.g. May be reliable as data from a reliable source (or based on large number of data)	B1 for a suitable comment about reliability.	

Question	Answ	er	Additional guidance	Mark
8 (a)	M1 A1	24.5 + 2×0.005 or 24.5 – 2×0.005 Warning lines drawn at 24.49 and 24.51	M1 for recognising that warning lines should be 2 standard deviations from the mean/target line. Implied by 24.51, 24.49, or one line correct (labelled or unlabelled) on diagram. A1 Both correct. Condone unlabelled if only 2 lines are drawn	(2)
(b)	B1	e.g. (Very) low values of sample range would indicate consistent components and suggest the process is under control/working well	B1 for a reason suggesting that low values of sample ranges are a good thing (in quality control) Condone e.g. 'range cannot be negative'	(1)
(c)	B1 B1	Means plotted at (1, 24.502) and (2, 24.496) on first chart Ranges plotted at (1, 0.006) and (2, 0.008) on second chart	B1 for both points correctly plotted B1 for both points correct plotted SC: If B0 B0 scored score B1B0 for any two correct points plotted out of 4	(2)
(d)	M1 M1 A1 B1ft	(24.498 + 24.492 + 24.501 + 24.509)/4 24.509 – 24.492 24.5 and 0.017 (3, '24.5') plotted on first chart and (3, '0.017') plotted on second chart	M1 for complete calculation of mean (implied by correct point plotted for mean) M1 for correct calculation of range (implied by correct point plotted for range) A1 for correct values (implied by both correct points plotted) B1ft for their points correctly plotted	(4)
(e)	B1ft	(First two samples are within all warning lines so process under control, but) Sample 3 has (sample mean on target but) range above action line so production line must be stopped/reset (oe)	B1ft for correct decision based on range of sample 3 (allow ft on calculation seen in part (d) or their plotted point) (If their range is between warning and action line, then allow e.g. 'take another sample/monitor the next sample' (If range is below action line, 'no action needed')	(1)

		NB: If no range stated or plotted, then B0.	
--	--	---	--

Question	Answer	Additional guidance	Mark
9 (a)	B1 More sales (in May 2019 than in May 2015)	B1 Assume comment is about 2019 unless otherwise indicated. Accept equivalent statements. Accept converse if clear reference to correct year is made.	(1)
(b)	M2 $30 \times \frac{90}{60} \times \left(\frac{4}{3}\right)^2$ or $30 \times 6 \times \left(\frac{4}{3}\right)^2 \div 4$	M2 for a complete method which uses ratio of angles and ratio of radii squared (allow their '3' and '4') (M1 for either ratio of angles $\frac{90}{60}$ or× 6 ÷4 or ratio of radii squared $(\frac{4}{3})^2$ implied by 45 or 53.3 or answer of 60 or (total frequency in 2019 =) 320 or $\frac{4}{3} = \frac{\sqrt{f_2}}{\sqrt{f_1}}$ oe)	(3)
	OR e.g.	OR	
	M1 $\pi \times 3^2 \times \frac{60}{360} = 1.5\pi \text{ (or 4.71) cm}^2 \text{ represents 30 packs}$ M1 $\pi \times 4^2 \times \frac{90}{360} = 4\pi \text{ (or 12.57)} \Rightarrow \frac{4\pi}{1.5\pi} \times 30$	M1 for finding relationship between area and number of packs. (allow their '3')	
		M1 for complete method using both areas and both angles. (allow their '4') A1 Allow integer answer from their '3'	
	A1 = 80	and their '4' from 76 to 84 inclusive	

Question	Answer	Additional guidance	Mark
10	M1 106.4×17 (+) 102.0×8 (+ 107.5×32 +) (= 10741.5)	M1 for an attempt to use weightings with index numbers in a product with at least two correct products. (May be seen in the table)	(5)
		Implied by any two of:	
		1808.8, 816, 3440, 675, 2722.5, 1279.2	
		(allow consistent subtraction of 100 or division by 100)	
	DepM1 '10741.5' ÷ (17+8+32+6+25+12)	M1 (dependent on previous M1) for complete attempt at sum of products divided by sum of weightings.	
	A1 = $107.4(15)$ or $7.4(15\%)$	A1 for awrt 107.4 or awrt 7.4(%)	
		NB: 107.3 is the mean of the six index numbers (without weightings) and scores M0M0A0	
	A1ft ('107.4' < 107.9, so) Cost of living has risen less for the Jones family, o.e.	A1ft for a correct conclusion in context (ft their 107.4) (dependent on both M marks) This is NOT for a comment about whether they are financially better/worse off	
	B1 e.g. Not possible to conclude as we do not know about income.	B1 for concluding 'not possible' with an appropriate reason relating to income or spending	

Question	Answ	ver	Additional guidance	Mark
11 (a)	B1	Normal (distribution)	B1 for correct distribution	(1)
(b)	B2	Temperatures are higher in the summer by $\underline{10.5}$ °C OR Mean temperature in summer is $\underline{14}$ °C which is greater than mean temperature in winter which is $\underline{3.5}$ °C .	B2 for a complete comparison with correct figure/figures (accept equivalent or converse statements) (B1 for an incomplete response, e.g. 'higher in summer' with no/incorrect figures or statement of correct figures with no comparison)	(2)
(c)	M1	0 to 7 represents , mean ± 3 standard deviations	M1 for recognising that all data lies within 3 standard deviations of mean	(2)
	A1	$(7 \div 6 \text{ or } 3.5 \div 3 =) 1.16(66) = 1.2 \text{ to one decimal place}$	A1 awrt 1.2 (scores 2 out of 2)	
(d)	B1	Winter temperatures have greater spread, or area under each graph is the same (or is 1)	B1 for equivalent comment recognising that greater spread leads to lower height Allow converse statement.	(1)
(e)	B2	Sample means will have less spread than daily temperatures, so Carol is not correct.	B2 for a complete answer with correct reasoning. (B1 for an incomplete answer, e.g. correct reasoning with incorrect/no conclusion, or correct conclusion with an attempt at reasoning.)	(2)
(f)	B1	15.6 is 2 s.d. above the mean/1.6 represents 2 s.d.	B1 for recognising 2 s.d. from mean (may be implied by 2 nd B1)	(5)
	B1	0.025 (probability of temperature above 15.6 °C)	B1 for correct probability of tail (allow 0.02275 from calculator)	
	M1	'0.025' × '0.025'	M1 for p^2 with $0A1ft for awrt 0.0006 (allow awrt 0.0005$	
	A1ft A1ft	$= 0.000625 \ (< 0.001)$ Greta is correct	from calculator) A1ft (dependent on M1A1) for correct conclusion consistent with their 0.000625	

Modifications to the mark scheme for Modified Large Print (MLP) papers: 1STO 2H

Only mark scheme amendments are shown where the enlargement or modification of the paper requires a change in the mark scheme.

The following tolerances should be accepted on marking MLP papers, unless otherwise stated below:

Angles: ±50

Measurements of length: ±5 mm

PAPER: 1ST0_2H				
Question		Modification	Mark scheme notes	
1		Wording added 'Look at the table for Question 1 in the Data Booklet. It shows an incomplete table.' Table enlarged and turned vertical.		
2		Modified Question Wording added 'Look at the diagram for Question 2 in the Data Booklet. It shows a box plot.' The word 'below' removed and replaced with 'in the Data Booklet'. Diagram enlarged. Open headed arrow. The horizontal axis label has been moved to the left. Values in the table changed to land on grid lines: 182 to 180 166 to 165 26 to 25 Table enlarged. The word 'above' removed and replaced with 'in the Data Booklet.		
3		Wording added 'Look at the table for Question 3 in the Data Booklet.' The word 'following' removed. Wording added 'shown in the table in the Data Booklet.' Table enlarged.		
3	(d), (e) & (f)	Wording added 'Look at the diagram for Questions 3(d), (e) and (f) in the Data Booklet. It shows'. The wording 'is shown below.' has been removed. Diagram enlarged. Axes labels moved to above the vertical axis and to the left of the horizontal axis. Open headed arrows. Crosses changed to dots. Grid intervals changed to 5. Small squares removed.		
3	(e)	Wording added 'in the Data Booklet.		

PAPER: 1ST0_2H				
Question		Modification	Mark scheme notes	
4	(a)	Wording added 'Look at Table 1 and Table 2 for Question 4(a) in the Data Booklet. Tables turned vertical and enlarged. The wording 'the following table' removed and replaced with 'Table 1 in the Data Booklet.' The word 'following' removed. Wording added 'shown in Table 2 in the Data Booklet'.		
4	(b)	Wording added 'Look at the table for Question 4(b) in the Data Booklet.' The wording 'below' removed and replaced with 'for Question 4(b) in the Data Booklet.' Tables turned vertical and enlarged. Wording added 'There are three spaces to fill.'		
5		Wording added 'Look at the information for Question 5 in the Data Booklet.' Information moved to the Data Booklet.		
7		Modified Question Wording added 'Look at the diagram for Question 7 in the Data Booklet. It shows a histogram.' The wording 'below' removed and replaced with 'in the Data Booklet'. Diagram enlarged. Shading changed. Open headed arrows. Small squares removed. Axes labels moved to above the vertical axis and to the left of the horizontal axis. Grid intervals changed to 1 on the vertical axis. Values on histogram changed to sit on grid line. The 5 to 10 block moved down to 13. The -10 to 0 block moved up to 1.5		
8	(b)	The number '14' changed to '17' in the question as the values have been changed. Wording added 'Look at Diagram 1 and Diagram 2 for Question 8 in the Data Booklet.' Diagram enlarged. Open headed arrows. Axes labels moved to above the vertical axis and to the left of the horizontal axis. Diagram 1 grid intervals changed to 0.005 on the vertical axis		

PAPER: 1ST0_2H				
Question		Modification	Mark scheme notes	
8	(a)	Wording added 'shown in Diagram 1 in the Data Booklet.' The wording 'below' removed and replaced with 'on Diagram 2 in the Data Booklet.' Diagram enlarged. Open headed arrows. Axes labels moved to above the vertical axis and to the left of the horizontal axis. Diagram 2 grid intervals changed to 0.002 on the vertical axis.		
8	(b)	Modified Question Table enlarged. Values changed to sit on grid lines. 24.502 to 24.505 24.496 to 24.495		
8	(c)	Wording added 'shown in Diagram 1 and Diagram 2 in the Data Booklet.'		
8	(d)	Wording added 'shown in Diagram 1 and Diagram 2 in the Data Booklet.' Leeway needed		
9		Wording added 'Look at the diagram for Question 9 in the Data Booklet. It shows two pie charts.' Diagram enlarged. Angle arcs made smaller. Shading changed. Key moved to above the diagram.		
10		Wording added 'Look at the table for Question 10 in the Data Booklet.' Wording added 'in the Data Booklet.' Table enlarged.		
11		Wording added 'Look at the diagram for Question 11 in the Data Booklet.' The word 'below' removed and replaced with 'in the Data Booklet' Diagram enlarged. Small markings removed. Interval labels changed to 1, with intermediate markings at 0.5. The horizontal axis label has been moved to the left. Dashed lines have been made thicker and longer. Open headed arrows.		

