



Pearson

Mark Scheme (Results)

November 2020

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

Publications Code 1ST0_2H_2011_MS

All the material in this publication is copyright

© Pearson Education Ltd 2020

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)	B2 Fully correct frequency polygon	B2 for all six points correctly plotted and joined with straight lines. Condone lines joining to 'x'-axis but not joining start to end. (B1 for six points plotted correctly (not joined), OR for at least three points correct and joined with straight lines)	(2)
(b)	<p>Any two from:</p> <p>B1 Correct comparison of average e.g.</p> <ul style="list-style-type: none"> • the number of children living in the houses was greater in 1918 on average • 1918 mode (4) > 2018 mode (1) <p>B1 Correct comparison of dispersion e.g.</p> <ul style="list-style-type: none"> • 1918 has a greater spread of number of children living in the houses/extends past 5 (children) • 1918 range (8) > 2018 range (5) <p>B1 Correct comparison of skew e.g.</p> <ul style="list-style-type: none"> • 2018 positive skew (but 1918 is more symmetrical) 	<p>Any two from:</p> <p>B1 for correct comparison of average.</p> <p>B1 for a correct comparison of dispersion. May refer to spread/range/variation.</p> <p>B1 for a correct comparison of skew.</p>	(2)

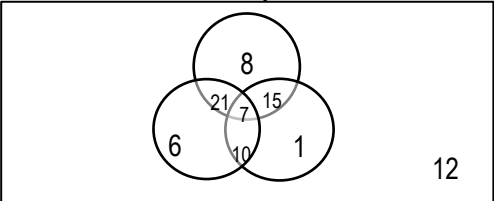
Question number	Answer	Additional guidance	Mark
2(a)	B1 Annual profit is the response variable because... <ul style="list-style-type: none"> • It depends on the distance to the car park • It is plotted on the y-axis 	B1 for a correct reason. Allow equivalent wording. Condone 'vertical' axis	(1)
(b)	B2 Mike is correct as the scatter graph shows negative correlation	B2 for a correct conclusion and mention of negative correlation. Allow a description of negative correlation provided it does not simply restate the question (B1 for correct conclusion with attempt at reason)	(2)
(c)(i)	B1 Straight line with intercept 40 000 B1 Straight line through (325, 27 000)		(2)
(c)(ii)	B1 (£40000 is) the profit when the distance (from the car park) is 0 (metres)	B1 for correct interpretation of intercept 40 000 in context	(1)
(d)	B2 Restaurant A/250m estimate is more reliable because... <ul style="list-style-type: none"> • involves interpolation • 250 is inside the range of data 	B2 for conclusion that A is more reliable (or B less reliable) with correct reasoning. (B1 for correct conclusion with attempt at reason)	(2)
(e)	B1 Conclusion is not valid because correlation does not imply causation	B1 for correct conclusion of not valid with a correct supporting reason	(1)
(f)	B1 64%	Allow awrt 64% Do not allow -64%	(1)

Question number	Answer	Additional guidance	Mark
3(a)	M1 for correct labelling of a scale A1 for 8 and 2	M1 implied by 8 or 2	(2)
(b)	M1 for correctly plotting one bar using their scale A1 for both bars correct on histogram (25 and 10)		(2)
(c)	B1 for positive (skew) B1 either correct interpretation <ul style="list-style-type: none"> • (the heights of trees) above the median have a greater spread • the mean (tree height) is greater than the median (tree height) 	B1 for correct identification of skew Do not allow positive correlation B1 for a correct interpretation of skew	(2)
(d)	M1M1 $\frac{1 \times 40 + 4 \times 120 + 32 \times 200 + 13 \times 280}{50} (= 211.2)$ A1 for David is incorrect with 211.2 B1 for identification of limitation of conclusion e.g. <ul style="list-style-type: none"> • Difference may not be the same at other locations • Both means are in the same class interval so we cannot be sure • We don't know the original data only the class intervals 	M1 for consistent use of fx with x within interval M1 for correct use of fx with x the mid-interval value with division by 50 A1 for correct mean and conclusion that David is incorrect/No B1 for identification of limitation of conclusion For the final bullet point do not allow 'these are just estimates' on its own.	(4)

Question number	Answer	Additional guidance	Mark
4(a)	B1 for e.g. <ul style="list-style-type: none"> • Not appropriate, collecting data at the local car park will give a unrepresentative/biased sample B1 for e.g. <ul style="list-style-type: none"> • Appropriate, all ages represented/age groups don't overlap • Not appropriate, (Quota sample) may not be proportional to the population • Not appropriate, not random 	B1 for a correct comment assessing the appropriateness of the location of the sampling B1 for a correct comment assessing the appropriateness of the method for sampling Do not allow small sample size	(2)
(b)	B1 for e.g. roll a dice / spin a spinner B1 for e.g. if the number is odd then answer the question 'do you drive faster than the speed limit on the motorway?' B1 for e.g. if the number is even then tick yes	B1 for an instruction to generate a random number B1 for indicating to answer the question for some outcomes B1 for indicating to tick yes for other outcomes	(3)

Question number	Answer	Additional guidance	Mark
5(a)	<p>B1 e.g. use heads to represent left, tails to indicate right</p> <p>B1 e.g. flip coin until get 4 heads or 4 tails in a row</p>	<p>B1 for indicating using one outcome for left and one for right</p> <p>B1 for description of simulation until 4 consecutive outcomes are the same</p>	(2)
(b)	<p>B1 B1 for any two from:</p> <ul style="list-style-type: none"> • the plan assumes all three drinks are equally popular/there should be 2 numbers for tea, 5 numbers for coffee and 1 number for hot chocolate • the simulation needs to be repeated multiple times to get an indication of variance • should only be using the numbers 1 to 8/There is no outcome assigned to the number 9 <p>B1dep The plan for simulation is not appropriate</p>	<p>B1 for each correct assessment (maximum 2) of the suitability of the plan</p> <p>B1 for identifying that the plan for simulation is not appropriate, dependent on B1 scored.</p>	(3)

Question number	Answer	Additional guidance	Mark
6(a)	<p>B1 e.g. people aged 0–24 make up a greater percentage of the population in India than in the UK</p> <p>B1 e.g. from 0–24 there are more males than females in both countries</p>	<p>B1 for correct comparison of overall percentages / proportion of the population in 0-24 age group Do not allow comments about numbers for this mark Ignore references to individual bars B1 for correct comparison of male/female division of the population in 0-24 age group Allow approximately the same number of males and females in the UK</p>	(2)
(b)	<p>M1 $2.8 + 3.2 + 3.3 + 3.3 + 3.1 + 3.2 + 3.6 (= 22.5)$</p> <p>M1 $\frac{22.5}{100} \times 65\,648\,000$</p> <p>M1 $\frac{775\,300}{14\,770\,800} \times 1000$</p> <p>A1 52.5</p>	<p>M1 for method to calculate total percentage of women in the 15–49 age range (may be implied by 22.1 – 22.9) M1 for method to calculate number of women in the 15–49 age range M1 for method to calculate general fertility rate A1 for awrt 52 or awrt 53</p>	(4)

Question number	Answer	Additional guidance	Mark
7(a)	B1 for 21, 10 and 8 in correct place B1 for 1 and 6 in correct place B1 for 12 in correct place 		(3)
(b)	M1 $\frac{15+7}{8+21+7+15}$ M1 $\frac{10+1}{"6"+10+"1"+"12"}$ A1ft $\frac{22}{51}$ or $\frac{11}{29}$ A1ft for $\frac{22}{51}$ (0.431) > $\frac{11}{29}$ (0.379) B1ft Inge is correct (oe)	M1 for correct method for conditional probability M1 for correct method for conditional probability. A1ft for awrt 0.43 or awrt 0.38 Allow ft from their Venn diagram in part (a) A1ft for correct comparison of conditional probabilities The final A1 is dependent on both M1s B1ft for correct conclusion based upon correct comparison of conditional probabilities	(5)
(c)(i)	B1 $\frac{22}{440}$ or $\frac{71}{710}$ B1 $\frac{71}{710} = 2$ $\frac{22}{440}$	B1 for finding an absolute risk Allow reciprocals B1 for showing relative risk is 2 (correct complete calculation) Allow any equivalent calculation leading to an answer of 2	(2)
(c)(ii)	B1 for e.g. A second hand car without a full service history sold from the garage is twice as likely to breakdown in the first year than a second hand car with a full service history sold from the garage.	B1 for correct interpretation of relative risk in context	(1)

Question number	Answer	Additional guidance	Mark
8(a)	B1 for both negative (correlations) B1 for Spearman's rank correlation coefficient closer to - 1/stronger	B1 must refer to both B1 allow 'Spearman's is lower' if first B1 scored	(2)
(b)	<p>B1 B1 B1 for any three from:</p> <p>Correlation:</p> <ul style="list-style-type: none">• Both have positive correlation• The correlation between wheat yield) and barley yield is stronger than between wheat yield and oat yield• Wheat is a better predictor of barley yield than it is of oat yield• Wheat yield and barley yield fit a linear model better than wheat yield and oat yield <p>Regression Equations:</p> <p>B1 for any one from:</p> <ul style="list-style-type: none">• Gradient of regression equation is greater for wheat yield and oats yield than for wheat yield and barley yield• It is meaningless to compare y –intercepts as you cannot have negative yields <p>B2 for e.g. Each additional t/ha for wheat would suggest an additional 1.52t/ha for oats / Each additional t/ha for wheat would suggest an additional 1.24t/ha for barley</p>	<p>B1 for each correct comparison (maximum 3) based on the correlation</p> <p>B1 for correct comment on either gradient or y-intercept</p> <p>B2 for contextual interpretation of gradient of regression equation (B1 for incomplete interpretation e.g. 'As wheat yield increases, both oat yield and barley yield increase')</p>	(6)

(c)(i)	<p>M1 $1.24 \times \frac{75}{28} - 0.30$ or $1.52 \times \frac{75}{28} - 1.05$ A1 for demonstrating both sides of the equation give the same value.</p> <p>OR</p> <p>M1 $-0.30 + 1.05 = 1.52x - 1.24x$ A1 $\frac{75}{28}$</p>	<p>M1 for substituting into one half of the equation given A1 for demonstrating both sides of the equation give the same value.</p> <p>OR</p> <p>M1 for a method to solve the equation formed A1 given answer from correct working</p>	(2)
(c)(ii)	<p>B2 e.g. if there is less than 2.67.. t/ha of wheat produced then plant barley rather than oats to get the larger yield (B1 e.g. if there is 2.67 t/ha of wheat then barley and oats would be expected to give the same yield)</p>	<p>B2 for contextual interpretation of the point of intersection of the regression equations (B1 for use of line of line of best fit with 2.67...)</p>	(2)
(c)(iii)	<p>B1 For any one from</p> <ul style="list-style-type: none"> • the data collected related to farms in Australia / only fields in sample • data does not take into account different growing conditions/weather • lower correlation for oat yield implies less confidence in the prediction. 	<p>B1 for a correct limitation of the data</p>	(1)

Question number	Answer	Additional guidance	Mark
9(a)	M1 $200 + 3 \times 2.5$ A1 207.5	M1 for use of $\mu \pm 3\sigma$ A1 for 207.5	(2)
(b)	B1 e.g. the process would be stopped / reset more frequently	B1 for a correct interpretation of effect of changing upper action limit	(1)
(c)	B2 for indicating that the machine should be reset (oe) as the range is outside the upper action limit although the mean is outside the upper warning limit (but not action limit) (B1 for indicating that the machine should be reset as the range is outside the upper action limit (no consideration of mean) / for indicating that as mean outside the upper warning limit another sample should be taken (no consideration of range))	B2 for a single correct decision (reset/stop the machine) with full reasoning (B1 for a decision with incomplete reasoning or for two different but correct separate decisions reached for the mean and the range)	(2)

Question number	Answer	Additional guidance	Mark
10(a)	<p>B1 B1 for any two from Appropriate since...</p> <ul style="list-style-type: none"> • they do not overlap • they are exhaustive (since no earthworm is greater than 36 cm) • unequal class intervals are suitable if data is closely grouped in one section of the range of values and more spread out in others • as normally distributed data would be more closely clustered in the middle of the range of values <p>Not appropriate since...</p> <ul style="list-style-type: none"> • it would be better to have larger class intervals towards 0 and 40 and smaller class intervals in the middle of the data range • it would be better to collect the data before deciding on the class intervals to use 	<p>B1 for each assessment (maximum 2) of the appropriateness of Kate's class intervals</p>	(2)
(b)	<p>B1 B1 B1 B1 for any four from</p> <ul style="list-style-type: none"> • Histogram will allow you to see if the distribution is bell-shaped • Histogram requires continuous data • Calculation of frequency density is incorrect as class width should be $10.5 - 5.5 (= 5)$ • 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 	<p>B1 for each assessment (maximum 4) of the appropriateness of using a histogram and calculating means and standard deviations</p> <p>Condone comments which refer to additional calculations that Bien could include to determine whether the data was normally distributed:</p> <ul style="list-style-type: none"> • Finding the median would also be appropriate • Calculating the skew would be appropriate 	(4)

Question number	Answer	Additional guidance	Mark
11	<p>M1 $\frac{4171}{39}$ (=107)</p> <p>M1 $\sqrt{\frac{469657}{39} - \left(\frac{4171}{39}\right)^2}$ (=24.585...)</p> <p>M1 $\frac{3("107"-97)}{"24.585..."}$</p> <p>A1 1.2139....</p> <p>B1ft for any one from</p> <ul style="list-style-type: none"> • the mean time is greater than the median time • the spread of times greater than the median time is greater than the spread of times less than the median time • the times are not normally distributed 	<p>M1 for method to calculate mean</p> <p>M1 for complete method to calculate of standard deviation</p> <p>M1 for using their mean and standard deviation in a complete method to calculate skew</p> <p>A1 for awrt 1.2</p> <p>B1ft for correct interpretation of skew in context</p> <p>Do not allow positive skew on its own for this mark</p>	(5)