Mark Scheme (Results)

November 2021

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

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

Question number	Answer	Additional guidance	Mark
1(a)	M1M1 $$\frac{2.5 \times 7 + 7.5 \times 10 + 12.5 \times 9 + 17.5 \times 6 + 22.5 \times 3}{7 + 10 + 9 + 6 + 3}$$ $(= 10.78 \dots)$ A1 for 10.8	M1 for consistent use of $\sum fx$ with x within interval (including ends) M1 for correct use of $\sum fx$ with x the mid-interval value (can be implied by 377.5) A1 for awrt 10.8	(3)
(b)	B1B1 for each of two limitations of conclusion e.g. • The data is just for Ben's office • Estimated mean in part (a) as we don't know the actual distances travelled/ have used midpoints • Don't know how the data for the newspaper article was collected/secondary data • Small sample size • Newspaper refers to 'average' – we don't *know* that this is the mean • We only know the average for 2001 and the year that Ben carried out his research.	B1B1 for each of two limitations of conclusion Do not accept Newspaper is out of date Source is unreliable	(2)

Question number	Answer	Additional guidance	Mark
2(a)	B1 Suitable as does not require specialist knowledge to understand/easy to understand. e.g. suitable because you don't have to be a statistics lecturer to understand the diagram. OR B1 Not suitable as it is difficult to distinguish between the colours.	B1 for assessment of the appropriateness of using a choropleth map	**(1)**
(b)	B3 for yes/partially AND two correct statements, e.g. - as shows countries in the south have higher literacy rates (above 60%) as they have more darker areas. - whereas countries in the centre/north have lower literacy rates (below 60%) as they have more lighter areas. - However, a small number of countries in the very north have higher literacy rates (above 60%) too. OR if B3 not earned… B2 for yes/partially AND one correct statement making reference to percentages or colours OR for two correct statements with no conclusion OR if B2 not earned… B1 for yes/partially with an attempt at a reason and not making reference to percentages or colours OR for a correct statement without a conclusion	B3 for a complete answer assessing the appropriateness of the conclusion OR if B3 not earned… B2 for an incomplete answer assessing the appropriateness of the conclusion. OR if B2 not earned… B1 for an attempt at assessing the appropriateness of the conclusion	**(3)**

Question number	Answer	Additional guidance	Mark
3(a)	B1 for negative (correlation) B1 for correct interpretation of negative correlation e.g. as the engine size increases the number of miles per gallon decreases	B1 for correct identification of correlation B1 for correct interpretation of correlation	**(2)**
(b)(i)	B1 19.39 ... B1ft for plotting (3.3, 19.4)	B1 for mean city fuel usage (implied by correctly plotting of the double mean point) B1ft for correct plotting of double mean point (3.3,'19.4')	**(2)**
(b)(ii)	B1ft for acceptable line of best fit through the double mean point	B1ft for acceptable line of best fit through the double mean point, should extend from 1.5 to 6 for engine size and pass through (3.3,'19.4') or their mean point ft their mean point SC if B0B0 scored in (b)(i) then B1 can be scored for an appropriate line of best fit	**(1)**
(c)	B1 for correct explanation of why this is not appropriate e.g. • Extrapolation / trend may not continue • beyond the range of values	B1 for correct explanation of why this is not appropriate	**(1)**

Question number	Answer	Additional guidance	Mark
4	B1 for each of five correct comments eg • the hypothesis should be a statement not a question • pre-test a good idea as it can identify issues/problems with the reaction time test • people might do better on the caffeine test because they have done the test before – biased • need to give more detail on test protocol e.g. how much coffee to drink, how long to wait after drinking coffee before doing second test • need to plan to control extraneous variables e.g. what time of day • people may not record data in spreadsheet accurately / may lie about their time • median/interquartile range are suitable measures to compare (as data may be skewed/have outliers)/needed to draw a box plot • good idea to consider outliers/may not be appropriate to remove outliers as they could be genuine data • box plots are suitable to e.g. show the distribution of the times • need to draw comparative box plots and not just one for after the coffee has been drunk	B1 for each correct comment on the appropriateness of the hypothesis or plans for collecting, processing and presenting the data	**(5)**

Question number	Answer	Additional guidance	Mark
5(a)	B2 for a correct comparison of the mean and standard deviation in context OR if B2 not earned… B1 Male tiger mean is greater than female tiger mean so male tigers have wider front pads (on average) B1 Male tiger standard deviation is greater than female tiger standard deviation so male tigers have more variable front pad widths	B2 for correct comparison of mean and standard deviation in context SC B1 for a correct comparison of the mean and standard deviation not in context	(2)
(b)	B1 Median more appropriate than mean when data is skewed / when there are extreme values/outliers/anomalies	B1 for correct comment identifying when the median is more appropriate than mean.	(1)

Question number	Answer	Additional guidance	Mark
6(a)	B1 for each of three correct comments eg - neither sampling method is random (or reference to one of the methods) /not everyone has an equal chance of being included - cluster sampling will not give a representative sample / only the views of a small number of branches/everyone in same restaurant might have the same opinion - judgement sample may be biased as the head chef may select people who will say positive things / agree with him - Only asking people who work kitchen staff who work at the restaurants so are likely to support the restaurant - judgement sample would mean that the people asked will have the experience to be able to answer the questions e.g. no one would be included who has only just started	B1 for each correct comment on the appropriateness of the sampling approaches	(3)
(b)	B1 $\frac{6}{28}$	B1 for correct probability Allow awrt 0.21	(1)
(c)	M1A1 Lines at 16, 38, 50, 78 horizontally and correct shading.	M1 for attempt at drawing a composite (i.e. stacked) chart, blocks in correct order OR at least 3 out of the 4 lines correct A1 fully correct percentage composite bar chart	(2)

(d)	B2 e.g. (New menu divides opinions) more because there are a greater proportion of 5(Excellent) ratings and a greater proportion of 1(Poor) ratings (than before) OR if B2 not earned… B1 for an incomplete answer e.g. greater proportion of 5 ratings (than before) OR greater proportion of 1 ratings (than before)	B2 for a complete interpretation of the percentage composite bar charts OR if B2 not earned… B1 for an incomplete interpretation of the percentage composite bar charts Condone reference to numbers rather than percentages	(2)

Question number	Answer	Additional guidance	Mark
7(a)	B1 e.g. Cost of computer storage reduces over time.	B1 for a suitable hypothesis including cost and time (B0 for a question)	**(1)**
(b)	M1 for either $\frac{24.87}{29.00} \times 100$ $(= 85.75 ...)$ OR $\frac{24.33}{24.87} \times 100$ $(= 98.2 ...)$ A1 for awrt 85.8 and awrt 98.2	M1 for correct calculation of chain-based index number. May be implied by one correct answer. A1 both correct	**(2)**
(c)(i)	M1 for $\sqrt[4]{93.8 \times 92.8 \times '85.8' \times '98.2'}$ A1ft awrt 92.5	M1 for correct calculation of the geometric mean of the four chain-based index numbers A1ft correct answer ft their (a)	**(2)**
(c)(ii)	B1ft for (average) 'decrease' …. B1ft …(at a rate) of '7.5'% <u>per year</u>	B1ft for correct contextual interpretation which must follow through their answer to c(i) B1ft for complete correct contextual interpretation of their value for geometric mean.	**(2)**

Question number	Answer	Additional guidance	Mark
8(a)	M1 for $\frac{114}{n} = \frac{69}{243}$ OR $\frac{114 \times 243}{69}$ oe A1 for 401 or 402	Must be an integer answer	(2)
(b)	B1 for reference to large sample size so it may be reliable B1 for reference to not knowing how the data was collected so may not be reliable	B1 for each of two correct reasons assessing the reliability of using this data Do not accept listing of individual assumptions of technique.	(2)
(c)	B1 answer in (a) is an overestimate/cause an increase in the estimate of the population size OR true population estimate would be lower oe		(1)

Question number	Answer	Additional guidance	Mark
9(a)	M1 $\frac{26}{48+26+22} \times 30$ (= 8.125) A1 8 or 9		(2)
(b)	B1 for a suitable question e.g. do you work hard consistently? B1 for an instruction to generate a random outcome e.g. roll a dice / spin a spinner B2 for a complete method to respond to the random response question e.g. if the dice roll is a 1 or 2 tick yes, if the dice roll is a 3 or 4 tick no, if the dice roll is a 5 or 6 answer the question OR if B2 is not earned… B1 for indicating to answer the question for **some** outcomes e.g. if the number is odd then answer the question		(4)

Question number	Answer	Additional guidance	Mark
10(a)	B2 Floor was the best performance (relative to other competitors) as it has the highest standardised score AND Pommel/rings are worst performance (relative to other competitors) as they have lower standardised score OR if B2 not earned… B1 for an incomplete answer e.g. floor was the best performance (relative to other competitors) as it had the highest standardised score OR pommel was the worst performance (relative to other competitors) as it had the lowest standardised score OR performance in floor was better than performance (relative to other competitors) in pommel as the standardised score was higher OR performance in floor was better than in pommel and/or in rings	B2 for a complete conclusion in context making reference to standardised score values. Must compare performance between three pieces of apparatus. Accept reference to floor being a positive standardised score /above mean (condone average) and pommel/rings being a negative standardised score /below mean (condone average) OR if B2 not earned… B1 for an incomplete conclusion	(2)
(b)	B1 eg - the apparatus have different distributions / means / standard deviations - scores in one apparatus may tend to be higher than in another - difficulty may be different for each apparatus	B1 for recognising that values are from different distributions making standardised scores appropriate	(1)
(c)	M1 $\frac{x-14.389}{0.854} = 0.247$ A1 14.6	M1 for forming a correct equation to solve to find the score A1 for awrt 14.6	(2)

Question number	Answer	Additional guidance	Mark
11(a)	M1 for 5×1 or 20×0.45 A1 for 5 and 9	M1 implied by 5 or 9	**(2)**
(b)	M1 for $\frac{11}{10}$ $(= 1.1)$ or $\frac{16}{5}$ $(= 3.2)$ A1 for both bars correct on histogram	M1 for calculation of frequency density (may be implied by one additional bar of correct height) A1 for both bars correct	**(2)**
(c)	M1 $\frac{5459}{70}$ $(= 77.98 \ldots)$ M1 $\sqrt{\frac{436137}{70} - \left(\frac{5459}{70}\right)^2}$ $(= 12.19 \ldots)$ M1 $"77.98 \ldots" - 3 \times "12.19 \ldots"$ $(= 41.39 \ldots)$ or $"77.98..." + 3 \times "12.19 \ldots"$ $(= 114.57 \ldots)$ A1 41.4 and 114.6 B1ft it is possible for there to be an outlier as the lowest class interval on the histogram/table extends to 40 so there could be an outlier between 40 and 41.4	M1 for method to calculate mean M1 for method to calculate standard deviation M1 for method to calculate one outlier limit A1 for both outlier limits B1ft for correct conclusion based on their limits for outliers and comparison to the histogram/table	**(5)**
(d)	M1 $\frac{3(75.6 - 79)}{11.8}$ $(= -0.864 \ldots)$ A1 -0.86 B1 e.g. • the mean adult male blood glucose level is less than the median adult male blood glucose level • the blood glucose levels are not normally distributed	M1 for method to calculate skew A1 for awrt -0.86 B1 for correct interpretation of negative skew in context Condone e.g. the spread of blood glucose levels above the mean is less than the spread of blood glucose levels below the mean	**(3)**

Question number	Answer	Additional guidance	Mark
12(a)	B1 0.35, 0.88, 0.12 in correct places	B1 for correctly completing the tree diagram	**(1)**
(b)	M1 $\frac{0.09\times0.65}{0.09\times0.65+"0.35"\times"0.12"}$ $(=\frac{39}{67}=0.582\ldots)$ A1 0.58	M1 for correct method to calculate conditional probability A1 for awrt 0.58	**(2)**
(c)	B1 for any one of • Independent trials • Probability remains constant	B1 for correctly identifying a property required for the binomial distribution	**(1)**
(d)	M1 for 0.91^6 or $6\times0.91^5\times0.09$ or $15\times0.91^4\times0.09^2$ or $20\times0.91^3\times0.09^3$ or $15\times0.91^2\times0.09^4$ or $6\times0.91\times0.09^5$ or 0.09^6 M1 $1-6\times0.91^5\times0.09-0.91^6$ or $15\times0.91^4\times0.09^2+20\times0.91^3\times0.09^3+15\times0.91^2\times0.09^4+6\times0.91\times0.09^5+0.09^6$ A1 0.095(15…)	M1 for using one correct term in binomial expansion, allow use of e.g $^6C_2$ for 15 M1 for a correct method to find probability of two or more items that do not meet the required standard A1 for awrt 0.095	**(3)**

Question number	Answer	Additional guidance	Mark
13(a)	B1 e.g. the time between successive calls is the same (at 27˚C) for both species of tree frog	B1 for contextual interpretation of the point of intersection of the regression equations	(1)
(b)	B2 e.g as temperature increases time between calls decreases (B1 e.g. both have a negative gradient/correlation/relationship) B1 e.g. the time between calls decreases more (for each 1˚C increase) for European tree frog than for Italian tree frog B1 e.g. one additional ˚C for Italian tree frog would give a decrease of 5.87ms in time between calls OR e.g.one additional ˚C for European tree frog would give a decrease of 7.9ms in time between calls B1 e.g. for lower temperatures/below 27˚C the time between calls for European tree frogs is greater than the time between calls for Italian tree frogs	B2 for contextual interpretation of negative gradients (B1 for identifying both have negative gradients) B1 for contextual interpretation of comparison of gradient of regression equations B1 for contextual interpretation of one gradient for a regression equation B1 for contextual comparison of the time between calls above or below 27˚C Condone e.g. at 0˚C the time between calls for Italian tree frogs would be shorter/182.30ms OR at 0˚C the time between calls for European tree frogs would be longer/237.2ms	(5)
(c)	B1 for indicating that the two groups of researchers may have used different methodologies e.g • Collected on different day/time • Different area of data collection • Different sample sizes • Different method of data collection	B1 for a correct limitation of comparing the results of two different sets of research	(1)