Pearson Edexcel

Examiners' Report
Principal Examiner Feedback

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 candidates at: www.pearson.com/uk

November 2021
Publications Code 1ST0_2H_2111_ER

GCSE (9-1) Statistics -2ST0

Principal Examiner Feedback – Higher Paper 2

General comments

It was a much smaller cohort of candidates (59 candidates) for this exam as it was a November resit after the cancellation of exams in Summer 2021. Candidates were generally able to attempt the whole paper within the time allowed and responded well to the challenges within the paper. It was pleasing to see candidates performed well on questions requiring standard techniques such as mean from a grouped frequency table (Q1), calculating chain based index numbers (Q7), histograms (Q11) and stratified sampling (Q9).

There were some notably improved responses from previous papers where candidates made clear decisions about appropriateness of methodologies and conclusions and made good attempts at justifying their responses.

Calculating conditional probabilities and using the binomial distribution to calculate probabilities remain challenging topics to candidates at this tier.

Question 1

Candidates made a strong start to the paper in part (a) and were able to work out the mean from a grouped frequency table. Out of the candidates who achieved one or two marks, some divided the sum of fx by 5 (number of groups) and not 35 (total frequency) and others used a value within the class interval, which wasn't actually the midpoint value.

When giving two reasons why Ben's conclusion may not be reliable in part (b) the most common correct answers given were that the data was just from Ben's office and that it was a small sample size. Many candidates were too vague in their statements and talked about how the newspaper was unreliable. Writing about the newspaper being out of date was a popular incorrect statement.

Question 2

In part (a) most candidates were able to discuss whether the choropleth map was suitable for the readers of the article by giving a conclusion and a correct reason.

In part (b) the majority of candidates were able to correctly decide that the choropleth map supported the conclusion and achieved two marks by making a correct statement on the literacy rates in the south when referring to percentages or shading. Few candidates achieved full marks as they did not make a comparison with the shading or percentages of literacy rates in the centre/north of the map or mention that the far north has some countries with higher literacy rates. A minority of candidates gave a correct conclusion but did not refer to percentages or colours on the map.

Question 3

In part (a) almost all candidates were able to correctly identify the negative correlation and interpret it in context.

In part b(i) candidates needed to calculate the mean value for fuel economics and then plot this point on the scatter graph. Of the candidates who could calculate the mean value correctly a small minority incorrectly plotted it by reading the scale on the x axis incorrectly and plotted at 3.6 instead of 3.3. A number of candidates failed to score the mark in part (b)(ii) where they needed to draw the line of best fit through the mean point. A common mistake was to not extend the line from 1.5 to 6 for the engine size.

In part (c) candidates were expected to indicate whether or not it was appropriate to use the scatter diagram to predict the fuel economy of a car with an engine size of 7 litres. Candidates generally answered this well with a large proportion also using the correct statistical vocabulary of extrapolation. The most common incorrect answer was that there wasn't any data for 7 litres.

Question 4

This extended response question was well received by candidates with most achieving at least 3 marks. Candidates needed to discuss whether Anne's hypothesis and plans for collecting, processing and presenting data were appropriate. The most popular comments recognised that the hypothesis should be a statement and not a question and related to the outliers. Many candidates also mentioned that the time waited between tests should be specific and that people may not record their own data accurately. There was some confusion about what a pre-test is and why it would be done. Many candidates thought that a pre-test was the candidates taking the reaction test before drinking the coffee. Candidates were also able to identify that calculating the median and quartiles was appropriate but were unable to support this with a reason such as the need to draw a box plot.

Question 5

In part (a) candidates were asked to make a comparison in context using the given mean and standard deviation of male and female front pad widths. Few candidates achieved full marks on this question but most were able to achieve one mark. Many were able to compare the mean and standard deviation, although few give a contextual interpretation of these comparisons or just compared one of these measures with a contextual interpretation. There were also some candidates who were confused with the contextual interpretation of the standard deviation e.g. male tigers have a greater standard deviation which means they have wider front pads than females.

In part (b) most candidates understood that the median is more appropriate to use when there are extreme values and this was the most popular response. Few candidates stated that the median is better when the data is skewed. The most common incorrect response said that the mean was appropriate as it uses all of the data, even though the question explicitly asked them to explain why the median was more appropriate.

Question 6

In part (a) candidates needed to discuss whether the two sampling methods are suitable. Most candidates were able to score one or two marks on this question by making a correct comment about the appropriateness of the sampling approaches. The most common marks to be scored were for stating that cluster sampling would not give a representative sample or for judgement sampling; only asking people who work in the kitchen would mean that they are likely to support the restaurant. A number of candidates compared the two sampling methods and suggested which method was better rather than discuss the appropriateness of each method.

In part (c) the overwhelming majority of candidates drew an accurate composite bar chart with correct shading. Of those who did not score full marks, they knew what they were doing but made small careless errors like forgetting to shade a block or putting one of the lines in the wrong place.

The majority of candidates scored at least one mark in part (c) for stating that there was a greater proportion of 5 ratings or a greater proportion of 1 ratings rather than mentioning both. Candidates needed to be aware that 2 comparisons were necessary because 2 marks were available here.

Question 7

Generally, candidates were able to suggest a suitable hypothesis in part (a) with only a few not scoring due to writing a question.

Calculation of a chain base index number, in part (b) of this question, was generally done correctly by candidates. Where incorrect answers were seen these were when they used 2015 as the base year and calculated the index number.

It was pleasing to see that a number of candidates were able to calculate the geometric mean of the chain based index numbers in part (c)(i). Where incorrect answers were seen the candidates worked out the arithmetic mean or when working out the geometric mean they added the numbers instead of multiplying the numbers.

The majority of candidates were able to identify that the geometric mean of 92.5 related to a 7.5 % decrease in part (ii), although some did not go on to give a full correct answer as they omitted the 'per year' from their interpretation.

Question 8

Most candidates were generally able to calculate and estimate for the total number of iguanas (part a) and give their final answer as an integer.

The marks available in part (b), discussing the reliability of using the biologists' data to work out an estimate, were more difficult to gain. Candidates instead listed the individual assumptions of the technique e.g. the tags could have fallen off. Very few wrote about the large sample size or about not knowing how the data was collected.

Nearly all candidates attempted part (c), but the answers given were sometimes a little confusing as to whether they were talking about the population estimate or the true population size. The most popular correct answer was talking about an increase in the estimate of the population size.

Question 9

Part (a) was a well answered question. Most candidates understood fully how to calculate the number required for a stratified sample. The most common answer was 8. The few errors seen were mainly due to careless errors in their working out.

In part (b) the majority of candidates scored one mark for a suitable question. For the candidates who did attempt a random response question, flipping a coin was by far the most popular method and most of these candidates understood how to describe the method for answering the question. Many of the candidates who mentioned using a die tended to miss out one or more of the outcomes and hence only indicated to answer the question for the outcomes mentioned.

Question 10

Nearly all candidates attempted part (a) to a varying degree of success. Many understood that the gymnast performed better on the floor compared to the other competitors on the rings or the pommel horse. Few candidates scored full marks due to not quite explaining enough about the standardised scores to back up their statement. It was common to see 'Brinn performed better in Floor Exercise as his score was closer to 1' or ‘Brinn performed worse in Pommel Horse as his score is closer to -1’ showing a lack of full understanding of the interpretation of standardised scores.

Few candidates obtained the marks in part (b) as they did not realise the difficulty may be different for each apparatus and hence they have different distributions.

It was pleasing to see in part (c) that the majority of candidates were familiar with the standardised score formula and were able to use it to achieve the correct answer. The minority of candidates who got this question incorrect could generally remember the standardised score formula, but then substituted the numbers into the wrong places or solved the subsequent equation incorrectly.

Question 11

Histograms with unequal class width was well understood and parts (a) and (b) were well answered.

In part (c) many candidates could use the given figures to work out the mean and standard deviation correctly. When working out the outliers however, some candidates got confused and tried to work out $1.5 \times$ IQR instead of $3 \times$ standard deviation. In addition, once the candidates had worked out the upper and lower outlier limits, few candidates explained clearly that it was possible there was an outlier in the lowest class interval or that the outlier may be between 40 and 41.4.

The vast majority of candidates could use the formula given on the formula page correctly to calculate the skew. However, the final mark was often not scored as the candidates could not interpret the skew in context and instead just stated that it was a negative skew.

Question 12

The vast majority of candidates scored the mark for part (a) and correctly placed the probabilities on the tree diagram.

Very few candidates were able to recognise or deal with the conditional probability in part (b). The most common error was to just do a calculation along the tree to find the probability of an item made by manufacturer A and state that the item does not meet the required standard (0.65×0.09). Those that did attempt to use the conditional probability formula often failed to use the correct probabilities in the formula.

In part (c) candidates were asked to write down one condition that needs to be assumed for a binomial distribution to be a suitable model. Many candidates incorrectly stated a fixed number of trials or only 2 outcomes but these were two conditions that were already given in the question. The most popular correct answer was that the probability of success remains constant.

It was pleasing to see nearly every candidate attempt part (d) of this question to varying degrees of success. Most candidates had an idea about the binomial distribution, but few achieved the final correct solution. Many could recognise that they needed to use 0.91 and 0.09 and raised them to the correct power, but then did not know which parts to use in the final solution or missed out the coefficients in their calculations.

Question 13

Part (a) was well answered and many candidates could explain in context that the times between calls were the same for both frogs at that temperature.

Though most candidates were able to access one or two marks in part (b), it was extremely rare for candidates to progress past 4 marks. The 2 marks often came from a correct interpretation of the negative gradients or from a correct comparison of the time between calls above or below 27°C, although candidates often made incorrect statements about the gradient e.g. European frogs has a larger gradient.

Candidates then attempted to compare the regression equations by making vague comments such as 'the regression equation for European frogs is higher than the one for Italian frogs'. Only the most able candidates described the effect of each additional °C, causing a decrease of 5.87m/s for Italian frogs or 7.9m/s for European frogs between calls. Candidates needed to consider the number of marks available as an indication of the amount of detail that was required in an answer.

The final part of this question was well answered and most candidates were able to give one potential limitation. The most popular comment was about the different methods of data collection.

Summary

Based on the general performance in this paper, candidates are offered the following advice:

- Ensure that correct statistical language is used throughout when making comparisons and conclusions.
- Develop understanding on comparing linear regression lines.
- Develop skills in assessing the reliability of using given data.
- Practise calculating probabilities using the binomial distribution.
- Practise calculating conditional probabilities.
- Develop skills in evaluating proposed sampling methods.

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom