Pearson Edexcel

Examiners' Report
Principal Examiner Feedback

November 2021

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

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

November 2021
Publications Code 1ST0_1H_2111_ER

GCSE (9-1) Statistics – 1ST0
Principal Examiner Feedback – Higher Paper 1

Introduction

General comments

Most students were able to respond to questions requiring completion of calculations or working with diagrams well and demonstrated good statistical understanding when asked to interpret these. Students were also generally able to make sensible comments relating to data collection approaches. Students found questions requiring evaluation of approaches or techniques slightly more challenging.

Students should be encouraged to show full working and set this out clearly so that partial credit can be awarded if a fully correct solution is not obtained.

Question 1

In part (a) the vast majority of students were able to give a reason why use of a pilot study is appropriate for the investigation described. Students often referred to checking that the questions were understood, but there were a variety of correct reasons given.

When asked whether or not it would be appropriate to carry out the pilot study on the entire population (part b) most students were able to identify a reason for this. A common answer was reference to time or cost which was condoned. Some students gave responses which had insufficient detail such as 'no – it is just a test run for a questionnaire' or 'it would not be appropriate as it defeats the aim of a pilot study'.

In part (c) most students were able to identify that a histogram would not be suitable for the data indicated and often referred to histograms being used for quantitative data. A minority of students incorrectly said that a histogram would be appropriate.

Question 2

Almost all students were able to correctly estimate the 75th percentile of the information shown in the cumulative frequency graph (part a(i)). It was noted that some students did not read the information in the question or use the total frequency shown on the cumulative frequency diagram and were observed to be working with a total frequency of 50 rather than 48.

Students found the interpretation of the 75th percentile (part (a)(ii)) more challenging. Common errors included giving an answer which did not give their interpretation in context or identifying that 75% of the counties in England have an area of 3400 sq km (omitting the 'or less').

In part (b) of the question students were told that half of the counties in England have an area between 2000 square kilometres and *k* square kilometres and asked to find an estimate for the value of *k*. This proved challenging for the majority of students with only a minority being able to find an estimated value for *k*. Some students were able to identify the cumulative frequency associated with an area of 2000 square kilometres making partial progress with the calculation.

Question 3

Part (a) of this question required students to identify the population for the two investigations. There were a significant proportion of fully or partially correct answers. A common error was to omit reference to the population including <u>all</u> of the students or <u>all</u> of the types of film or identifying UK cinemas as the population in (ii).

The majority of students were above to identify suitable advice to give to Susan to ensure that the information she collects from the internet is reliable (part b). Many students indicated using reliable websites, ensuring that the data was up-to-date or checking the information collected using another source.

In part (c) of the question students were asked to give a reason why each of the two sampling methods described were not random. Method A was quota sampling and a significant number of students indicated that the students in each year group did not have the same chance of being selected. A common incorrect answer for method A was indicating that this was not random because

John was manually selecting the students which was not the case. Method B was opportunity sampling and there were a significant number of students who indicated that the students selected would be those that arrived together, would be likely to be in friendship groups or not everyone had an equal chance of being selected.

Part (d) of the question asked students to explain which of the two methods should be used to minimise bias. It was pleasing to see that a significant proportion of students were able to identify that method A would be more likely to be representative. Some students correctly identified method A, but were not able to give a correct reason.

Question 4

This question was well answered by the majority of students. In part (a) most students were able to correctly identify the year and quarter that the UK came out of the recession.

Students were generally able to calculate the simple index number for the gross domestic product of the UK in 2008 Quarter 1. Where incorrect answers were seen this was generally due to having the numerator and denominator in the calculation interchanged.

In part (c) students were required to work backwards from the simple index number and gross domestic product for the base year to find the gross domestic product for 2010 Quarter 1. This was also well answered.
The final part of the question asked a student to comment on a given conclusion. The majority of students were able to indicate that the conclusion given was incorrect and explain that the index numbers increased.

Question 5

In part (a) the vast majority of students were able to correctly identify the quarter each year from 2014 to 2017 where the recycling was between 42% and 44% based upon the time series graph.

Part (b) of the question asked students to describe the seasonal trend shown by the time series graph. Many of the students were able to correctly identify seasonal trends from the time series graph. Some students commented on

quarters 2 & 3 being above the trend line each year and quarters 1 & 4 being below the trend line each year which was awarded partial credit. The most common incorrect error was to identify the overall trend shown which was not what was required.

In part (c) students were asked to identify the number of points which would be appropriate to use in a moving average to help describe the trend and give a reason for their answer. Students were generally able to indicate that 4-point moving averages would be appropriate and that the data was given in quarters. Common incorrect answers included 44 which was from attempting to find an estimated average from the graph.

Many students were able to correctly indicate that use of the time series graph to make the prediction suggested in (d) would be unreliable and it was common to see reference to extrapolation for the reason. Some students incorrectly stated that this would be reliable and referred to a clear trend.

Question 6

This was an unusual question on Venn diagrams and proved challenging for students.

In part (a) students were asked to identify which two events from *A*, *B* and *C* were mutually exclusive and give a reason for their answer. Whilst a reasonable number of students were able to identify that *A* and *C* were independent it was less common to see a correct reason. Common incorrect reasons appeared to be based on confusing mutually exclusive (as in the question) with independent. Other incorrect responses included students giving probabilities as their answers, commonly 0.38 and 0.05 or 0.08 and 0.23.

In part (b) some students were able to correctly calculate $P(B)$ based on the Venn diagram, however there were a significant proportion of students who gave the answer 0.15 having failed to realise that they needed to include the regions intersecting with *A* and with *C*.

It was pleasing to see that a significant proportion of students were able to correctly find $P(A \text{ or } C)$ based upon the values in the Venn diagram. Where incorrect responses were see these were generally from omitting some of the

values in their calculation. It was notable that some students had a correct calculation process, but made errors in evaluating their calculation.
Part (d) requires students to complete a Venn diagram for *A* and *B* based on the original Venn diagram. A minority of students were able to give a fully correct Venn diagram as their answer. There were a significant number of students who were able to correctly place the values for 0.08 and 0.38, but then gave 0.43 rather than 0.2 (including the probability for only *C* with the probability for *B* not *A* or gave 0.11 as $P((A \cup B)')$.

Question 7

Part (a) of the question was a fairly standard question asking students to give an advantage of using comparative pie charts rather than using ordinary pie charts. A significant proportion of students were able to indicate that the comparative pie charts would allow for comparison of relative frequencies. The most common incorrect response was to refer to being able to compare proportions which is not a correct advantage of comparative pie charts compared to ordinary pie charts as both allow for comparison of proportions.

In part (b) of the question students were asked to make comparisons of the angles of the sector representing the British museum in the comparative pie charts for 2005 and for 2018. They were also asked to compare the areas for the sectors in the pie charts for the British museum in the comparative pie charts for 2005 and 2018. Many students were able to give a correct comparison for at least one of the two features, but correct comparisons for both was less common. In some cases students performed calculations in order to make the comparison, however this was not required.

Part (c) of the question required students to calculate the radius of the comparative pie chart for 2018. This is a standard requirement in a question on comparative pie charts. Only a minority of students were able to correctly calculate the required radius.
The most common error was to calculate $\frac{12.5}{8.3} \times 7.5$

Question 8

In this question students needed to use the information provided to calculate limits for outliers, identify outliers from the section of spreadsheet provided and draw a box plot with outliers marked. Students were generally able to draw a correct box plot for the data, but many did not attempt to calculate outliers. Where students did attempt the outlier calculations then these were not always successful. There were a minority of students who were able to give a fully correct box plot with outliers.

Question 9

This question was answered well by a pleasing number of students. The most common method used was to assume that Anders scored 100 in the fourth assignment, calculate a weighted mean and identify that this was greater than 98. Where the answer was not fully correct it was common to see students using weighting with the three assignment marks given in the question. A common incorrect approach was working with mean average rather than a weighted mean.

Question 10

Part (a) of this question required students to identify the need to calculate Spearman Rank Correlation Coefficient and use this to comment on the hypothesis given. It was pleasing to see a significant proportion of students attempting the calculation of Spearman Rank Correlation Coefficient, although there were a surprising number of students who made errors in the ranking. There were also a significant number of students who did not attempt to calculate the correlation coefficient and instead commented on the hypothesis based upon the data values.

Part (b) of the question asked how the reliability of the conclusion in (a) could be improved. A majority of students were able to give a correct answer often commenting on collecting more data.

Question 11

This question tested student understanding of quality control charts and how the distribution of the mean volume of 4 cans would compare to the distribution of the volume of the individual cans.

Only a minority of students were able to draw a fully correct quality control chart. Common errors were use of 4 rather than 2 as the standard deviation, omitting the scale or spacing the warning and action lines such that the distance from the target volume to the warning line was the same as the distance between the warning and action lines.

In part (b) a pleasing number of students were able to correctly explain that the actions were not appropriate as the machine should have been stopped and reset the first time the action limit was exceeded. A common error was to identify that the actions were appropriate.

Part (c) of the question asked students to comment on the claim that the volume of orange juice in individual cans should be modelled by the same normal distribution as the mean volume per can of orange juice in a sample of 4 cans. Only a minority of students attempted an answer to this part of this question and fewer still were able to make correct comments relating to the means and standard deviations.

Question 12

Almost all students were able to correctly identify the outlier on the scatter diagram (part a), but fewer students were able to give a correct comment on the price of the car represented by this point (part b). A common incorrect response in part b was to give the age of car and price represented by the point or to describe the trend shown on the scatter diagram.

Part (c) asked students to give a reason why it might be appropriate to remove the outlier and a reason why it might not be appropriate to remove the outlier. There were a significant number of correct reasons given for removing the outlier and for keeping the outlier.

In part (d) students were asked how the value for Spearman Rank Correlation Coefficient would compare with the value of Pearson's product moment correlation coefficient. Students were expected to identify that the correlation coefficient would be closer to -1 and that this is because the pattern is not linear. There were a minority of fully correct answers showing clear understanding of how the correlation coefficients compare. There were however a good number of students who were able to indicate that the value for Spearman Rank Correlation Coefficient would be negative. Common incorrect responses were based on an interpretation of the correlation given.

Question 13

Part (a) of this question was answered well by students with the majority able to find the 2.5^{th} to 97.5^{th} interpercentile range.

In part (b) students were asked to calculate a probability based on the percentile data provided in the table. This proved more challenging for students.

Part (c) required students to give a reason why using the mean and standard deviation of the weights of the cats to summarise the data was appropriate. There were a small number of excellent answers to the question, however the majority of students struggled with the question and there were also a significant number of students who did not attempt to answer.

Summary

Based on their performance on this paper, students should:

- Practise writing clear explanations, bearing in mind exactly what is asked in the question and what evidence you should give to support your answer.
- Practise interpreting statistical calculations in the context of the question.
- Develop skills in evaluating approaches to statistical methods.
- Practise calculating the radius of comparative pie charts and develop understanding of the usage of comparative pie charts.
- Develop understanding of how the distribution of the set of sample means compares to the distribution of individual values from the same population.
- Practise drawing and using quality control charts.
- Practise calculation of Spearman's rank correlation coefficient.

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