Pearson Edexcel

Examiners' Report
Principal Examiner Feedback

Summer 2019

Pearson Edexcel GCSE (9 – 1)

In Statistics (1ST0)

Foundation Paper 1F

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

Grade Boundaries

Grade boundaries for all papers can be found on the website at:
https://qualifications.pearson.com/en/support/support-topics/results-certification/grade-boundaries.html

Summer 2019
Publications Code 1ST0_1F_1906_ER

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

Introduction

General comments

Students were able generally able to attempt the whole paper within the time allowed. Students performed well on questions requiring standard techniques such as completing charts (2a, 3b, 9ab, 11a), interpreting charts (2b, 3c, 6bc, 10a) and calculation of median or range (4ab).
Students were able to attempt questions assessing the new style of assessment (AO3), but they would benefit from more practice on this type of question as the quality of response was variable.

Question 1

This question required students to demonstrate understanding of likelihood (parts a and b) and the probability scale (parts c and d). Students were generally able to answer all parts of the question correctly. Where incorrect answers were seen this was generally indicating 'likely' in part (a) and 'blue and red' in part (b). Very few incorrect answers were seen to parts (c) and (d).

Question 2

This question required students to complete and interpret a bar chart. In part (a) students were asked to complete the bar chart using the information provided. The majority of students were able to draw a bar of height 3. Where incorrect responses were seen this was often drawing a bar of height 4. In part (b) students needed to use the bar chart to identify how many more people gave a 1 star rating than a 3 star rating, again this was generally answered well with the majority of students able to give a correct answer.
In part (c) students were asked to identify why a conclusion might not be validity based upon the bar chart. The majority of students were able to give a clear justification for this with many referring to the number of ratings other than 5-star being greater than the number of 5-star ratings, there being lots of 1 and 2-star ratings, or referring to the number of 1-star ratings and 5-star ratings being similar. Some students referred to the number of 1-star ratings, for example saying '5 gave 1-star', but this was not enough on its own – there needed to be a comparison to either the total number of ratings '5 gave 1-star out of 20 overall' or to the number of 5-star ratings '5 gave 1-star which was

similar to the number of 5-star'. A minority of students made reference to the small sample size of only 20 ratings, but this did not use the bar chart to answer the question.

Question 3

Parts a to c and e of this question were answered well by students.
In part (a) of this question students were asked to state what type of data would be found in a newspaper. Many students were able to correctly identify that this would be secondary data, although an aspect of secondary data, for example 'unreliable' would also have been accepted but was rarely seen. Some students incorrectly stated quantitative or discrete, although the data in the question could be described using these statistical words this did not answer the question posed which specifically required the type of data that would be obtained from a newspaper.
Part (b) involved students completing the tally chart and frequency column. The majority of students were able to produce a fully correct tally chart with correct frequencies. Where errors were made these were often due to incorrect tallying of one or more of the values given – often the overall total was correct. A minority of students did not group their tallies in fives, but were often able to gain a mark for a correct row or frequency column. Students should be encouraged to check totals.
In part (c) students were required to determine the modal group. The majority were able to correctly identify the group '3 or 4'. Where incorrect answers were seen these included giving the mode of the original data or giving the frequency of the modal group.
Part (d) proved more challenging for students. Although a good number of fully correct answers, or correct answers following through the students' frequencies, were seen it was also common to see incorrect answers. Some students added too few or too many frequencies, with an answer of 27 from adding the first four frequencies being relatively common. A small number of students worked out $2010 - 1981$ using the years mentioned in the question.
In part (e) students were asked to identify a suitable diagram to display the data. The majority of students were able to give an appropriate diagram, with bar chart being the most common response. Some students indicated diagrams which were not suitable for the discrete data collected, for example histograms. Another common incorrect response was scatter diagram.

Question 4

Parts (a) and (b) of this question were answered well, part (c) proved more difficult for students.
In part (a) students were asked to calculate the range of the data. The majority of students were able to do this correctly. Where students could not correctly calculate the range some gained 1 mark for identifying both the highest and lowest values, often

giving the answer 4.3 to 68.2, or, more frequently, for a subtraction with one value correct. Incorrect answers included calculating the median and calculating the mean.
In part (b) students were asked to find the median of the data. It was pleasing to see a significant proportion of fully correct answers to this as students needed to order the data and correctly find a median from a list with an even number of pieces of data. The majority of students knew that they needed to list the values in order. Some missed a value from the list and a significant proportion did not know how to find the median where there were an even number of pieces of data and gave the answer 9.2, 10.6 or 9.2 and 10.6. Some students struggled to calculate the middle value between 9.2 and 10.6. A minority of students incorrectly calculated the range or the mean.
Part (c) required students to explain why the median was a more appropriate average than the mean for the set of data provided. It was pleasing that some students had clear understanding of the affect of extreme values on the mean and these students were able to justify the choice. There were a significant proportion of incorrect answers seen to this part, with some students incorrectly referring to one being an average and the other not, some referring to simplicity of calculation and others explaining what a median is.

Question 5

Part (a) of this question was not answered well by students. A minority of responses were fully correct. Very few recognised the data is not bivariate data and for a scatter diagram to be used the data needs to be in related pair form. A greater proportion of students indicated that the diagram was not appropriate and attempted a reason, with reference to difficulties of showing male and female on the same scatter diagram being a popular attempt at a reason. Some students stated that the diagram was not suitable but made no attempt at a reason or indicated that an alternative diagram would be better, these responses were not awarded a mark. A significant number incorrectly believed that a scatter diagram was appropriate for the data with some indicating that they believed this was the case because the data was bivariate (which it was not).
Part (b) of the question was answered well. Many students were able to correctly identify that it would be difficult to access the tennis players. Where incorrect responses were seen a common error was to write about the suitability or not of using secondary data / data from the internet.

Question 6

In part (a) of this question students were asked to design a data collection sheet that could be used. There were a high proportion of students who drew suitable data collection sheets. Some students only gave a partially correct answer by giving a column for takeaway or, more commonly, listing takeaway options, but not indicating tallies/frequencies/number for the second column. A small number of students produced a data collection sheet that was too generic in that they had a tally and/or

frequency column but did not label a column ‘takeaway’ nor give any takeaway options. A common incorrect response was to give a question for a questionnaire, another common incorrect answer was a sketch of a bar chart.
Parts (b) and (c) of the question were generally correctly answered with the majority of students able to correctly interpret the pie charts to answer the question.
In part (d) students were asked to identify whether or not the two pie charts supported the statement that the number of male customers ordering meat is the same as the number of female customers ordering meat. Students needed to recognise that the pie charts show proportions rather than amounts and so cannot be compared in this way unless we know that the totals are the same. There were a good number of fully correct answers to the question with students generally indicating that we did not know the numbers of males and females (rather than referring to pie charts representing proportions). A small number of students were aware of comparative pie charts (which is not part of the foundation specification) and indicated that the pie charts could be used to support the claim because both were the same size and the meat segments are equal, this was condoned. Common incorrect answers indicated that the pie charts could be used to support the claim because both had a quarter for meat.

Question 7

In part (a) of this question students were required to select examples of types of data from the list given. The majority were able to identify type of game as an example of categorical data. Part (a)(i) caused slightly more difficulty with a number of students incorrectly indicating that average playing time was discrete data.
Part (b) of the question asked students to describe how to take a systematic sample. A minority of students were able to give a fully correct answer. It was more common to see partial descriptions of systematic sampling, often omitting the random start point. Common incorrect answers were descriptions of other sampling methods, with a description of random sampling being a common incorrect response.

Question 8

Part (a) of this question was answered well with many students indicating that Suresh would not be allowed to access the records. Another acceptable response which was frequently seen was an indication that the records might be out of date or may not record absences. Some students made reference to the data being personal or confidential, but this alone was not sufficient as the question asked for a problem with the data collection and these are not problems with the data collection (although they may be the cause of problems). Incorrect responses often made reference to reasons for absences or gave generic responses such as ‘time consuming’.
In part (b) of the question students were offered two possible methods for calculating a mean from a frequency table only one of which was correct. They were asked to identify which was the correct method and explain why. Only a minority of students

gave a fully correct response, indicating Tia and giving a complete explanation of why this was the appropriate calculation – division of total number of absences (multiplication of number of days by frequency) and the use of the total number of people (division by 30). It was more common to see a partial assessment of the appropriate choice – Tia together with one of the two elements – this was awarded 1 mark. There were a very significant proportion of incorrect answers where students indicated Ami.
In part (c) students were required to work out the median from a frequency table. Some students were able to successfully calculate this using $(n + 1)/2$ to identify that the 15.5th value was required and then finding this. Others took the longer, but valid, approach of listing all the values out and crossing off from each end until they reached the middle, sometimes correctly identifying that this was the 15th and 16th values and sometimes just being left with one of the two (leading to 1 mark being awarded for method). The majority of students obtained an incorrect answer, this was commonly from finding the median of the frequencies (3) or the median of the number of days options in the table (4).

Question 9

Students were generally able to interpret the scatter diagram to find the number of passengers for the individual ship as required in part (a).
Part (b) required students to draw a line of best fit on the scatter diagram, the majority of responses to this were correct. Where students did not have an acceptable line of best fit then this was either because it failed to follow the trend of the data (generally too steep) or due to the candidate joining the individual plots in a chain (treating this as a dot to dot).
Part (c) required students to describe and interpret the correlation shown on the scatter diagram. The three elements required were 'positive', 'strong' and a contextual interpretation. The majority of students were able to gain at least one mark for this part, often for stating positive or for indicating that as the weight increases the number of passengers increases. A significant number of students were awarded 2 marks for indicating that it was positive correlation and giving a contextual interpretation. Only a minority of students gave a full description of the correlation and interpreted in context.
In part (d) students were expected to indicate whether or not it was appropriate to use the scatter diagram to predict a value in the future. Some very good responses were seen with students often indicating that it was not appropriate and that this was outside the range of data. A significant number of students were awarded 1 mark, generally for identifying that this was not an appropriate thing to do together with an attempt at a reason, commonly that the line of best fit may be inaccurate. There were also a significant number of students who incorrectly indicated that this was appropriate often citing the strong correlation as justification for their answer, not appreciating that trends can change once you have gone beyond the data given.

Question 10

In part (a) students were expected to read information from the composite percentage bar chart and perform an appropriate calculation. The majority of students were able to correctly identify the required percentage. Where incorrect answers were seen this was often due to errors in reading the scale.
Part (b) of the question required students to demonstrate understanding that the percentage composite bar charts showed percentages and not the numbers of listeners. Correct answers were less common than incorrect answers. Common incorrect responses included comments on the structure of the bar, for example indicating that the categories were stacked. Some students incorrectly suggested some of the category was not included in the diagram or that the diagram was misleading. Other incorrect answers included those where the candidate indicated that the conclusion was correct event though the question stated it was not.

Question 11

In part (a) of this question students were asked to complete a choropleth chart. The majority of students were able to do this successfully. Incorrect responses included using the incorrect shading for some of the squares or, in a minority of cases, shading squares with a combination of two sections of the key.
In part (b) students were required to comment on the validity of a conclusion based upon the choropleth diagram. It was common to see a conclusion of 'valid' together with reference to more children in that area. Where students did not give both a decision and a reason using the choropleth map, they were not awarded the mark. Other incorrect answers generally did not make reference to the choropleth map for their supporting reason.

Question 12

In part (a) of this question students were expected to find the lower quartile, median and upper quartile from a cumulative frequency curve. Many students were able to do this correctly. Where incorrect answers were seen these were sometimes based on using 110 rather than 100 as the total frequency or incorrectly working from the horizontal axis up to the curve and then across to the vertical axis.
In part (b) the majority of students were able to make an attempt at a box and whisker plot, generally with at least two correct plots (often the lowest and highest value) which was awarded 1 mark. There were a good number of fully correct box plots or box plots which were correct following through from the students' values in part (a). Some students struggled with the scale on the box plot and therefore made errors in plotting.
Part (c) required students to make comparisons from the two box plots and interpret one of their comparisons. Correct comparisons of the median, interquartile range or range were reasonably common, with many students also able to include a correct

interpretation. Comparisons of skew were far less common, but were generally correct when attempted. A significant number of students gave comparisons of lowest, lower quartile, upper quartile, highest which did not gain marks. In a minority of cases students listed the values for median, IQR or range, but did not make comparisons.

Question 13

In part (a) of this question students were asked to calculate the number of female students that should be in a stratified sample. There were a good number of fully correct answers. Most students that calculated correctly did realise to round the number to whole number of people. Where students gave incorrect answers, these were commonly 100 (selecting equal numbers of male students and female students) or from dividing the number of female students by 200. There were a significant number of students who did not attempt this part of the question.

Part (b) asked students to give a situation in which it would not be suitable to stratify by gender. A common correct answer was to refer to an investigation which only relates to one gender, but a small number of students also referred to a situation in which a comparison was being made between the genders. Incorrect answers included reference to situations where it would be appropriate to stratify by gender or consideration of elements other than a statistical situation, for example ethical or moral considerations.

In part (c) students were asked to identify an alternative method of stratification for this data. Many students were able to correctly indicate type of school. Where incorrect answers were seen these included references to types of sampling, age (not included in the data) or reference to only one or two of the types of school.

In part (d) students were asked why two events were not mutually exclusive. Some good answers were seen, the most common correct response being 'you can be female and go to an independent school'. There were, however, a majority of students who could not answer this question correctly. There were a range of incorrect responses, with students confusing mutually exclusive events with independent events, for example indicating that 'they don't affect each other'. Other incorrect responses included referring to males being able to go to independent schools too, seemingly interpreting the question as asking why independent schools were not exclusively for girls.

Part (e) of the question asked students to find a probability. There were a large number of students who did not attempt this question. Where students did attempt the question some did give the correct probability, but others gave the probability of just one of the two and some gave the answer 9351, failing to recognise that this was not a probability.

Question 14

Part (a) of this question was not well answered. Students were asked to indicate why including interquartile ranges in a newspaper article about how house prices have changed over time would be appropriate. Very few realised it is a comparison of the spread of the middle 50% and in this case whether this variability has increased or decreased between the two years. This mark generally was only achieved if students mentioned how the figure would not be affected by extreme values. Many gave answers which were too vague, for example 'because you would have more data'. Others incorrectly referred to comparing averages or made a comment which suggested that interquartile range would give a measure of the change in house prices (rather than the variability of house prices).
Part (b) of this question was answered correctly by a greater proportion of students than part (a). Many students demonstrated awareness that the audience might not understand interquartile range. Incorrect answers varied quite widely, with references to the interquartile range not being interesting to the audience or the information harming house sales.

Question 15

This question proved challenging for students with a significant number of completely incorrect answers. A reasonable number of students were able to gain the first two marks for calculating some of the values that were needed to make the comparison, but either not attempting a probability or attempting probabilities with incorrect denominators (commonly 40), a large number of students attempted to use a probability tree diagram but they gave incorrect probabilities. The most successful students were those that drew a two-way table to calculate the values. A minority of students gave fully correct responses.

Question 16

Part (a) of this question required students to calculate the crude birth rate using the formula that had been provided in the question. The majority of students were able to do this correctly. Incorrect responses included examples of correct substitution followed by incorrect evaluation (gaining 1 mark) or students who changed the 1000 in the formula to 100. A minority of students did not attempt this part of the question.
Part (b) of this question had a reasonable number of fully correct answers where students indicated that the two populations would need to be equal in size. Common incorrect answers included reference to one population being twice the other (seen both ways around) or showing that one of the birth rates was twice the other by giving the calculation $12 \times 2 = 24$.

Question 17

This question required students to interpret and compare values of gradient and intercept on the y-axis for the line of best fit for price plotted against mileage for two different models of car. A number of students just compared the numerical values for gradient and intercept without interpretation, as the question required contextual comparisons this was not sufficient. Where students made correct contextual comments these commonly included reference to both cars decreasing in price as their mileage increases, model A's mileage affecting the price more than model B's or that model B had a greater price. An incorrect contextual comparison that was seen several times was that the higher the mileage, the more expensive the car. Some students also incorrectly interpreted intercept and gradient as referring to higher or lower mileages for the cars. Students need to recognise that for five marks usually five different statements need to be given.

Summary

Based on their performance on this paper, students should:

- read each question fully and carefully before attempting to answer it
- show working out to support the final answer
- practice interpreting answers to statistical calculations and diagrams in context
- practice identifying the appropriateness of calculations, diagrams and approaches in different settings
- give a decision when the question asks for this and support it using the information from the question (in the case of conclusions) or properties of the statistical techniques (when commenting on appropriateness of techniques)
- practice on making comparisons, particularly those in context (12c, 17).

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