Pearson Edexcel

Examiners' Report

Principal Examiner Feedback

November 2020

Pearson Edexcel GCSE (9-1)

In Statistics (1ST0)

Foundation Paper 2F

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

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

November 2020
Publications Code 1ST0_2F_2011_ER

GCSE (9-1) Statistics -1ST0

Principal Examiner Feedback – Foundation Paper 2

Introduction

General comments

It was a much smaller cohort of students (90 candidates) for this exam as it was a November resit after the cancellation of exams in Summer 2020. Students were generally able to attempt the whole paper within the time allowed. It was pleasing to see students performed well on questions requiring standard techniques such as probability (Q1), interpreting of secondary data, both from diagrams (Bar line graph in Q2 and population pyramid Q3) and calculation of median and range (Q5).

It was also very pleasing to see students attempt the extended response question but like in 2019 they would benefit from more practice on this type of question as the quality of responses was variable.

Like 2019, students were still less familiar with some topics new to Foundation tier (Q12, Risk).

Question 1

This question required students to demonstrate understanding of the probability scale (part a) and likelihood (part b). In part (a) students showed reasonable understanding of the probability scale, but found it more difficult to create their own event to show a particular probability for the letter B ($\frac{1}{6}$). Correct answers seen where often the events ' a number 1' or 'a number 6' . Incorrect answers seen here were answers where often events with more than one number e.g. 'a number 1 or 2' and a small minority of students left this part of the question blank,

perhaps not even seeing that they had to fill in an event. Very few incorrect answers were seen in part (b).

Question 2

This question required students to interpret a bar line graph. In part (a) the students were asked to read from the graph, and the majority of students were able to read off the correct month. Very few incorrect responses were seen here. In part (b) students needed to select the correct word to complete the sentence. Students were not as successful on this part of the question, with a large amount of incorrect answers given with students selecting one of the three incorrect words. In part (c) students were asked to determine whether more than half the months in 2017 were wet months. The majority of students were able to clearly state that more than half the months were wet months by stating that 7 of them were wet months or there were only 5 'not wet' months although a small minority failed to make clear what their conclusion was. Some students gave the correct conclusion but failed to give any supporting figures so were not awarded any marks for this part of the question.

Question 3

In this question students were given a population pyramid where they were asked to read figures from it and compare age groups. In parts (a), (b) and (c) the majority were able to correctly read the answers from the population pyramid with very few incorrect answers given. Part (d) was well answered with most students able to identify that the percentages were the same for male and female. Some students gave the correct figure of 12.7%, although this was not needed to gain the mark. Some students confused the number of people with percentages and said that the number of males and females were equal. In this particular

population pyramid this was awarded the mark as it was a population pyramid for the total population of Africa. In part (e) students were asked to give a reason why the sum of the percentages is 100.1%. Students were not as successful on this part of the question and many failed to mention that the figures had been rounded. Incorrect answers included mistakes when counting, people being born after the percentages were calculated, and outliers in the data. A large number of students left this part of the question blank.

Question 4

In this question students were given secondary data for car colours in the UK and a pictogram showing the data. Part (a) was well answered and students were able to explain why the data was secondary, usually stating that the data had come from the internet or the BBC. Part (b) was also well answered and students correctly completed the key for the pictogram. In part (c) students needed to decide if the pictogram was the most appropriate way to display the data. Many students did comment that it was difficult to accurately represent the percentages of the car in the pictogram and that it was not appropriate. When incorrect answers were seen it was to make comments just stating that the pictogram was hard to read/understand without giving a reason why. Some students also failed to state that it was not the best diagram or that Jill was correct. Few chose to suggest that a pie chart would be a better way to represent the proportions or that the key only showed what full cars represented. In part (d) students who identified that blue was not the mode generally got full marks for giving one of the correct reasons on the mark scheme. Many students failed to read this question carefully and gave the incorrect answer that blue was the mode as it had the highest frequency as they were looking at Table 2 instead of Table 1. Part (c) required students to explain why the mode was a more appropriate average to use for the information in Table 2. This part of the question was not answered

well and students failed to recognise that they couldn't work out any of the other averages or comment as to why, i.e. because the data was non-numeric. Instead, most students give the advantages of the mode e.g. not affected by extreme values, or just described how to find the mode or stated the definition of the mode. Part (f) required students to compare the number of silver cars in Tables 1 and 2. The majority of students incorrectly said that there were more silver cars in Table 1 as they hadn't taken into account that Table 1 was giving the percentages of all cars and Table 2 was the frequency of cars in a car park. Some students did gain one mark by stating there were more cars in Table 2 but failed to achieve the marks for calculating 10% of 60 and comparing this to the value given in Table 2.

Question 5

In part (a) of this question students had to calculate the longest time for Puzzle X using the range and the shortest time. The most common incorrect answer seen was to add the range to the median to get 40. In part (b) students were asked to use the information to compare the median time for X and Y and to interpret the comparison. It was pleasing to see a large proportion of students gaining both marks for this question. When marks were lost it was for just stating the medians and not comparing them and missing an interpretation in context. Part (c) of this question was well answered by all students. They were able to find the range of the times. Part (d) was less successful and a large proportion failed to understand that the word 'consistent' related to the range and many incorrect answers were seen here. Some stated that it was X because it had a bigger median or Y because it had a larger range. In part (e) students needed to explain what Hannah had done wrong when calculating her median from the stem and leaf diagram. It was pleasing to see that many students were awarded full marks on this question as they identified that Hannah had failed to order the stem and leaf diagram and

then found the correct median of 37. Where incorrect responses were seen a common error was to give the median value as 7 and not take into account the stem or missing out a value when they drew an ordered stem and leaf.

Question 6

This question required students to calculate the crude birth rate and number of births using the formula that had been provided with the question. The majority of students were able to calculate the birth rate correctly. Incorrect responses included examples of students who misread the 1000 in the formula for 100 or incorrect substitution of the values given in the table by missing out numbers. They found using the formula to find the number of birth rates more of a challenge as they needed to rearrange the formula. Incorrect responses included not dividing the answer by 1000, not rounding the final answer to an integer (so losing the final accuracy mark) or just dividing the total population by the crude birth rate (giving an answer of 550,186). There was a large proportion of students who didn't attempt this second part of the question.

Question 7

Part (a) was successfully answered by most students. They were able to correctly comment on why a sample was more appropriate, with the most common answers including that it was quicker and easier or giving a converse statement that needed to include the word census to gain the mark. A common incorrect response was to give a disadvantage of taking a census and not to include the word census e.g. it takes longer. In part (b) students needed to describe one problem during the statistical enquiry process that may be encountered and explain how to overcome the problem. There were many vague answers or no description of what could lead to non-response or unexpected outcomes but

students were more successful in suggesting ways to overcome the problems. The most common explanation of what problems she may encounter were that students might not know how long they do chores for or that they don't do chores at home and the most common suggestions of how to overcome the problems where to include option boxes for them to tick. A significant number incorrectly said that the students may lie which was not an acceptable answer.

Question 8

In Part (a) students were asked to write down a suitable sampling frame. It was clear that 'sampling frame' was not commonly well understood, with students instead describing ways of taking a sample e.g. random sampling. In part (b) they needed to name the sampling method that was in the plan. There were very few correct answers and often students just wrote down any statistical word that they knew that wasn't a type of sampling method e.g. 'questionnaire' or 'categorical'. Part (c), worth 5 marks, was one of the extended responses to be expected on the new GCSE. Students needed to discuss whether the plans in sections A, B and C were appropriate. They needed to give a correct comment for each section with a maximum of 3 marks from any one section. It was pleasing to see that the majority of students were able to achieve some marks in this question but very few achieved all 5 marks. Students mainly picked up the mark when making comments about sections B. Students demonstrated knowledge of open and closed questions and that question 4 was a leading question. They also recognised that question 5 is not exhaustive and didn't include a time-frame. Less well expressed but often relevant were comments on section A. Where correct answers were seen it was often stating that the sample method was not random or it was appropriate to take students from every year group. Section C was where fewer marks were scored for this question. Many chose to comment on the

difficulty of producing the graphs rather than the appropriateness of the graphs chosen. Where marks were scored for this section, they were for recognising that the pictogram was appropriate or that they hadn't asked if the students were vegetarian to be able to draw the graph. In part (d) students needed to give an advantage and disadvantage of using a class width of 10p instead of 5p. Where correct answers were seen for the advantages they stated that it would be quicker to handle the data. For the disadvantages, correct comments recognised that the answer would be less accurate. Incorrect answers for the advantages included that it would be easier to understand and for the disadvantages that it would be inaccurate. Many students did not recognise that this question was asking about the suitability of the class widths and gave answers about money instead e.g. Mehmet may not have any 10p coins and only have 5p coins or that it may be difficult to collect 10p. Other incorrect disadvantages said that it would be time consuming to fill in a table or that the table didn't start at zero, again not understanding what the question was asking. Many students did not attempt this question.

Question 9

Part (a) was a more standard question and was well answered by the majority of students. They were able to calculate the mean and very few incorrect answers were seen. When incorrect answers were seen, students often still picked up the method mark for showing their working out. Students found part (b) of this question difficult as it was a question that needed careful reading. In conclusion 1 students needed to identify that Daniel referred to each of the months and that an average would not give the fact about all of the months was not seen. Most students focused on their answer in part (a) and stated that this was higher so was invalid. Some marks were awarded for conclusion 2 where students recognised that the comment wasn't valid as we only knew about the first 6

months of 2018. Where incorrect answers were seen again it was because students didn't read the question carefully and said that it was valid as the mean for 2018 was greater than 2017. Many students didn't attempt this question or didn't give a reason why the conclusions weren't valid.

Question 10

This scatter graph question was the first common question on the paper (also on 2H). Many students picked up marks in parts (b), (c) and (d). In part (a) students were asked to explain why annual profit is the response variable. Foundation students found this very difficult and had little or no knowledge of what a response variable was. Instead, students gave a variety of answers that had nothing to do with a response variable. In part (b) many students were able to pick up one mark by stating that it didn't support the hypothesis and make an attempt at a reason. While some correctly stated 'negative correlation', many tried to support it with vague descriptions or by simply repeating the question, which was not awarded the mark. This question says 'giving a statistical reason' which a large proportion of students failed to do. In part (c) students needed to draw a line of best fit using the mean point and intercept. The majority of students drew a line of best fit by eye on their scatter graph without any reference to the information given in the table and losing one or both marks. In part (ii) of this question students were asked to interpret the value of the intercept. This part of the question was often left blank and when students did attempt it they were unable to interpret the value. Incorrect responses often just wrote down the value of the intercept (40 000) or wrote down a description of the correlation shown on the graph. In part (d) of this question students needed to explain which of two estimates would be the most reliable. Many students were able to identify that A was the most reliable and make an attempt at a reason to be awarded one mark.

Few were able to support their decision with the correct answer of interpolation. The most common incorrect reasons given were A as it's closer to the car park or not B as there is no data for 700m or close to 700m, not fully explaining that this was extrapolation or out of the range of the data. In part (e) students needed to explain that the conclusion was not valid because correlation does not imply causation. Students failed to recognise this and gave the incorrect answer that the conclusion was valid as more tables meant more people and therefore more profit as it showed positive correlation. In part (f) students needed to calculate the percentage decrease for the top 100 restaurants. The majority of students failed to even give an answer for this question and left the answer line blank. When answers were given often students just wrote down random numbers or calculated 345-125 = 220 and then divided this by 10 or 100 to get a number less than 100.

Question 11

Question 11 was the second common question on the paper. In parts (a) and (b) students were given an incomplete histogram and incomplete grouped frequency table and were asked to complete both. Students could successfully complete the table, often without completing the scale on the histogram and were awarded full marks. Students often were awarded one mark for having one of the values correct. If students hadn't added a scale on the histogram they then often didn't complete the histogram correctly or at all. When the scale had been correctly labelled in multiples of five students then went on to complete the histogram bars correctly and were awarded full marks. Those choosing to work in other multiples, like 3, often ended up with incorrect graphs, but could have been awarded the method mark for part (b) for using their scale and plotting at least one bar. It was rare, but some students did attempt to try to work out frequency density as they had seen that it was a histogram. It is worth noting that calculating and using

frequency density to draw histograms (unequal class widths) is only on the Higher tier. In part (c) students needed to identify and interpret the type of skew shown on the histogram. A small proportion were able to identify it as a positive skew, but the incorrect answer of 'negative skew' was often given. Students were unable to interpret the skew and instead chose to describe the distribution in general terms rather than using statistical language. In part (d) students were given a grouped frequency table for two different types of trees and asked to show their working to decide if David was correct and give a limitation of their conclusion. Not many students were clear on the method for estimating the mean from grouped data and were awarded no marks. Those that did have some understanding of the method would often multiply by the wrong mid- interval or incorrectly divide their summation by 4 (number of groups) instead of 50. A number of students also went on to calculate the mean for Field Maple trees (which was already given in the question) and not for Silverleaf Maple trees, as a consequence of not reading the question carefully. Those that did successfully calculate the mean for Silverleaf Maple then often went on to make the correct conclusion that David was incorrect. A number of students gave a conclusion but showed no working and therefore were awarded no marks. The majority of students failed to read the question again after they had completed their calculations and conclusion and made no attempt to give a limitation. When a limitation was seen it was that the answer was an estimate and not one of the answers given on the mark scheme.

Question 12

Question 12 was the final question on the paper and was not a common question. Students were asked to calculate and interpret relative risk. This new topic of risk proved a challenge for the majority of students with a significant number not understanding relative risk. Very few students could calculate the relative risk in

part (a) and often instead subtracted the two values in the table of 245 and 35 to give the answer of 210. Where marks were awarded it was often for given one of the risk of sunburn wearing sunblock ($\frac{35}{400}$). In part (b) students needed to give an interpretation of their relative risk found in part (a). Students often gave the answer 'if you wear sunburn you are less likely to get sunburnt', as they knew this was the common sense answer, but had failed to attempt part (a) so were not awarded a mark for an interpretation of their relative risk. Again in part (c), students showed a lack of understanding of relative risk. When students did attempt this answer it was not to interpret the relative risk of 1 but more to make a comment about what they thought about drinking tap water and upset stomachs. The most common incorrect answer was "drinking tap water gives you an upset stomach".

Summary

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

- Practise the statistical techniques new to Foundation tier, including relative risk.
- Students still need to practise topics that they have struggled to understand previously e.g. definition of a sampling frame, names of sampling methods and response variables.
- Read each question fully and carefully before and after answering to ensure you haven't missed out any details. For example, Question 11(d) asked for a limitation and almost all students failed to attempt this and they also worked out the mean for the wrong tree. In question 4 (d) a large proportion of students gave answers relating to the wrong table.

- Give a decision when a question asks for it and use information or working out to support your decision. A number of these types of questions needed data to support their conclusions e.g. Question 2(c), Question 4 (f) and Question 11 (d).
- Ensure that correct statistical language is used throughout when making comparisons and conclusions.
- Attempt all questions and do not leave a question line blank.