edexcel

Principal Examiner Feedback

Summer 2016

Pearson Edexcel GCSE in Statistics (2ST01) 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

Summer 2016
Publications Code 5ST1F_01_1606_ER

GCSE Statistics 2ST01
Principal Examiner Feedback – Foundation Paper 1

Introduction

Students on the whole seemed to find the paper accessible and generally had time to attempt all questions. Students seemed to find this paper slightly more challenging than last year's paper, particularly Q11, Q13, Q14 and Q15. There are some topics, in particular, which were not well attempted, including understanding statistical terminology ('mutually exclusive' and 'closed questions'), using the multiplication rule to calculate a probability, calculating an estimate of a mean from a grouped frequency table and calculating moving averages.

Students should show their working as some may have picked up more credit when their answers were incorrect (Q1(b), Q4(b), Q6(b), Q13(c)). The standard of drawing diagrams was often quite pleasing although students must be reminded that when drawing a trend line on a time series graph, using a ruler is appropriate.

In Statistics, many responses rely on explanations and it was pleasing to see an improvement in the clarity of written expression though this remains an issue for some students at this level. With comparison and interpretation, especially where a question is indicated as QWC (marked with *), students should take extra care to write clearly and use correct statistical language. Where more than one mark is available for a question, students should be aware that the number of marks generally indicates the number of comments expected.

Report on individual questions

Question 1

Overall, this was a successful start to the paper by the majority of students. Virtually all students were able to read the key correctly in part (a) to work out the total number of votes for Edward. Most students went on to find correctly the total number of votes in part (b) though some slips cost students 2 marks when no working was shown. To score the mark in part (c) students needed to give both a supporting reason and a correct conclusion. The most common error was simply to state that 'Rita got more than half the votes', with no supporting figure.

Question 2

At Foundation Level most students are very familiar with probability likelihood terminology with most students scoring the marks in parts (a)(i) and (a)(ii). The most common mistake was to confuse 'likely' with 'certain' and 'impossible' with 'unlikely'. Again in part (b)(i), the vast majority of students indicated the position of Y correctly on the probability scale between ½ and 1. Students were less successful in part (b)(ii) where some students gave the answer 'likely' rather than the required numerical probability.

Question 3

Students found Q3 more demanding, particularly parts (c) and (d). In part (a), although the majority of students scored the mark for working out the correct number of days, others incorrectly included the frequency for 15°C. Another common mistake was to give an answer of 3 presumably thinking that there were 3 temperatures in the table that were less than 15°C. Most students were able to identify the mode as 14°C in part (b). Here the most common mistake was to give the frequency of 4 as the answer. Other students stated that both 1 and 3 were the mode as these frequencies appeared the most often.

Parts (c) and (d) required explanations and most found these parts challenging. Many students thought that 20.2°C could not be the mean because it had a decimal in it. Those who said that 20.2°C was not one of the numbers in the table did not make it sufficiently clear that the data set did not go higher than 17°C. Some students incorrectly gave the mean as 14.5°C neglecting to take into account the frequencies. Some students accurately but unnecessarily calculated the mean and were able to score the mark. The majority of students scored no marks in part (d) for incorrectly stating that the mode would be greater, usually by reasoning that the mode would be more appropriate than the mean or median. Some students stated that the modes would be equal without giving a supporting statement but this was rarer than gaining full marks.

Question 4

This question was accessible to all students with good performances seen in parts (a), (b) and (c). Most students selected ‘cereals’ from the table as the category with the highest mean amount of money spent each year. There were a number of students, however, who gave the year 2012 rather than the category. Part (b) was very well answered but many students did not heed the advice to show their working. Some students lost a mark here by confusing the units and coming up with an answer of 600p or 0.06p. Again, there was little difficulty in identifying the required categories in part (c).

Part (d) saw a variety of incorrect answers including ‘it was added up wrong’ and ‘there’s a 1p difference’. Only the most able students appreciated that these figures were rounded despite this fact being clearly stated in the table. Finally, in part (e) students were required to compare Sophie’s information with the relevant information from the table. Far too many students only compared Sophie’s figures with each other stating that she spent more eating at home than she did eating out. Those students who scored only 1 mark here often summed the figures to make an overall comparison of spending.

Question 5

Overall, students performed well on this question. Virtually all students gave the correct answer ‘vanilla’ as the most popular flavour of ice cream in part (a). Part (b) required a calculation to show that the number of people whose favourite flavour is strawberry is 4. Many students gave incomplete reasoning by saying that each sector represented 2 people, without explaining that there were two sectors for strawberry on the pie chart. Part (c) provided a good source of marks for all students as most students confidently transformed the data from one representation to another. Common mistakes were to misread the scale and to label the axes as ‘x’ and ‘y’.

Question 6

This question discriminated well. A surprisingly low proportion of students recognised that 50% of the data lies above the median in part (a). Though a good number of students were able to use the box plot to calculate the range of the data in part (b), far too many students opted to calculate the interquartile range instead. Other students struggled with the scale and gave 80 – 2 = 78 as a common incorrect answer. The description of the skew of the distribution in part (c) was handled only by the higher achieving students on this paper. Interestingly enough part (d) had an even lower success rate with most students unable to identify the median as the type of average used to compare box plots.

Question 7

Parts (a) and (b) were very well answered. It was pleasing to see students in part (c) stick to a description of the overall trend rather than giving a year to year description of the data. Part (d) saw mixed success. Those students who paid attention to detail realised that not all age groups were included on the graph. Other students did comment that the information was only from the UK. Incorrect answers here usually referred to the difficulty in counting all of the cinema admissions rather than referring to the graph. A number of students mistakenly believed that the graph did show this information for all of the years.

Question 8

Even though it was necessary to justify each answer, there were some pleasing performances from students on this question. Most students made a promising start with by writing down the correct age group in part (a).

For the first claim in part (b), some students misunderstood the scale and believed that the claim was actually true though a large proportion of students did work out that ‘it went up in 2s’. Other students looked at the wrong age group and said that the correct percentage was 20%. Very few students ticked false without making an attempt at a supporting reason.

Students found it more difficult to give a supporting reason for why the second claim was true. Most students just recopied the claim without making any reference to the multiple bar chart. It was not uncommon for students to believe that the claim was true since all of the bars were above 45%, perhaps being confused by the fact the scale only went up to 90%.

Many students correctly realised that the final claim was false as it was not true for those who were aged 17 – 20 and 21 – 29. Even those students who ticked true could score the mark here if they made it sufficiently clear that it was only true for 5 of the age groups.

Question 9

This question tested the students' understanding of sampling methods and questionnaires and was generally well attempted. Students gave a variety of suggestions for the population of the survey but many students failed to realise that it should include all members of the choir. At this level many students still believe that the population refers to the number of people in the group and '100' was a common suggestion.

Students had prepared well for part (b) with the standard 'textbook' answers of 'quicker' or 'easier' being fairly common. The most common error in part (b) was to re-use the already given 'cheaper'. Though a good proportion of students did select 'bias' as the best term from the list in part (c), 'time' and 'cost' were selected quite often.

Part (d) was very well answered as most students were able to distinguish between quantitative and qualitative data. It was part (e) that was the most demanding with 'open questions' being chosen for the survey by the majority of students. Of those that did opt for 'closed' most were unable to provide a valid reason why. It is clear that understanding the difference between open and closed questions is an area that needs improvement.

In part (f) 'leading' or 'biased' were suggested as the problem with the question by a good proportion of students. Many students, however, were concerned that there were only two options boxes available, whilst other students expressed a dislike of classical music.

Question 10

Despite the slightly unusual demands for a question on scatter diagrams, it was pleasing to see good performances by students here. Most students correctly identified that this was indeed a scatter diagram in part (a). Part (b) generally saw good success though on a number of occasions students switched the axes and incorrectly located the point for Republic of Ireland at (0.9, 2.4).

Most students had little trouble identifying the correct point in part (c) and circling it on the diagram. It was unfortunate that some students identified other points as well and therefore lost the mark by circling more than one point. In part (d), a pleasing proportion of students recognised that the points showed no correlation and that a line of best fit would not be appropriate.

Question 11

Whilst good attempts were made at the basic probabilities in part (a), parts (b) and (c) were beyond the level of most of the students on this paper. Of those students who wrote a numerical probability in part (a)(i), virtually all gave the correct one. As usual, weaker students tended to give their answers in words with 'likely' seen often. An equally high success rate was seen in part (a)(ii) with very few students missing seeing the word '**not**' and giving an answer of $\frac{2}{10}$.

It was clear that 'mutually exclusive' is a term unfamiliar to Foundation Level students and very few correct responses were seen. Some students attempted a definition of independent events whilst others suggested it meant something equivalent to random. Part (c) was equally challenging as the majority of students gave an answer of $\frac{3}{10}$ without going any further. Some students went on to multiply this by 2 but it was rare to see the probability multiplied by itself.

Question 12

This question on data collection proved testing at this level. In part (a) the students were generally able to provide a question to test the preference of cats/dogs. The most common error was the lack of response boxes.

In part (b), a neat, useable two-way data collection sheet was rarely seen. Most students scored 1 mark for a table which was suitable for recording gender. Common errors for students who scored no marks were to design a questionnaire or they simply drew an empty box.

Most answers to part (c) concerned the number of data points available rather than the type of data that was being collected. A common response was to state that scatter diagrams were for two variables and there were more than two types of animals.

Question 13

At this stage of the paper only the higher performing students were able to make much progress in showing a good understanding of grouped data. Part (a) was generally successfully attempted, although some wrote the frequency '22' rather than the class. Clarity of written expression varied in part (b) as many students seemed to recognise that the data were grouped but struggled to explain it convincingly. Some students thought the groups 'overlapped' and others were concerned that the presidents might still be growing.

Though a large number of students left part (c) blank, many of those who attempted part (c) recognised that they had to find the class midpoints. Many students, however, just added these midpoints and divided by 4 or by 43. There was also a lot of confusion as to what to do with the 43. Most incorrect answers were outside the range of the data and students should check that their answers make sense.

Those who understood what was being asked in part (d) generally opted for the 180.1 cm estimate as being the better of the two estimates. Again students found it hard to articulate a correct reason here with many saying 'the data in the second table is more spread out'.

Question 14

Most students found this question on sampling challenging and it was rare to see students scoring more than 2 out of 6 marks here. The majority of students failed to score the mark in (a) either by leaving it blank or by not including enough information to demonstrate fully how the stratum size of 15 can be determined.

Despite being provided with the random numbers in part (b), many students felt obliged to use a calculator to generate even more random numbers. A large number of responses included the use of a hat, which again was completely unnecessary. The best responses did discuss using a numbered list of students and matching the random numbers to the students to obtain the sample. Only a few students realised there was a repeated number (53) and that this needed to be ignored.

A significant proportion of students, who scored no marks in parts (a) and (b), still managed to score both marks in part (c). An incorrect answer of $\frac{42}{90}$ was common.

Question 15

Students are encouraged to try all questions on the paper since those students who persevered to the end often scored at least 1 or 2 marks here. It appears that many students had not prepared well for time series questions, particularly for the calculation of moving averages in part (a), as many students tried to identify a pattern to complete the table. Occasionally the previous 5 moving averages that were given were averaged rather than the appropriate figures from the absences column.

It was unfortunate that in many cases when an answer was given in part (a), there was no attempt to plot it on the time series graph, as there was an opportunity for students to gain a follow through mark here. Those who did plot the point in the correct region of the graph tended to do so accurately.

In part (c) students were expected to use a ruler to draw a single, straight trend line. There were a large number of attempts simply to join up each of the moving average points.

Finally for those who gave an answer to part (d), many students were rewarded with a mark for saying that the number of absences on Friday was increasing. Less common was to explain that there were more absences on Friday compared with the other days of the week.

Summary

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

- read each question fully and carefully before attempting to answer it and check that answers make sense in the context of the question
- show clear working to support the final answer and when necessary give a clear decision as well as the reasons
- write down probabilities as fractions, decimals or percentages and remember that a demand for a probability requires a numerical response, whilst a demand for likelihood requires a word response
- try to attempt all questions
- use a ruler to draw straight lines and a protractor for angles
- understand and use the correct terminology such as “open and closed questions” and “mutually exclusive events”

Grade Boundaries

Grade boundaries for this, and all other papers, can be found on the website on this link:
http://www.edexcel.com/iwantto/Pages/grade-boundaries.aspx

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