Examiners' Report

Principal Examiner Feedback

Summer 2017

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 2017

Publications Code 5ST1F_01_1706_ER

GCSE Statistics 5ST1F
Principal Examiner Feedback – Foundation Paper 1

General Comments

Students found this paper generally accessible, especially the beginning, although at the top end some questions were found to be a little more demanding than usual; examples of this were those who chose to draw a probability tree from scratch in Q15b or the contextual request for comparison in Q14c without the previously often seen demand to "compare distributions". They usually scored well on familiar topics such as pictograms, bar charts and two-way tables but rarely scored well on a demand for understanding more technical terminology such as sampling frame or skew (when asked to describe the shape of a distribution). Other low scoring parts included requests to choose appropriate diagrams or to describe details of the random sampling process.

Sometimes parts of questions were left blank; possibly suggesting that particular techniques or terminology were not known rather than any lack of sufficient time. There was some evidence of a lack of care in reading questions. This led to students having a false perception of what the question was asking for, or led to them making incorrect or irrelevant comparisons.

Where calculations are required students should be encouraged to show correct mathematical working. For example subtractions were needed in a number of places (eg Q6b, Q13a and Q14a) but when the two figures were written down with a minus sign they were often in the wrong order. This was only condoned if students then found the correct positive difference.

Report on individual questions

Question 1

This question provided a very familiar and accessible start to the paper for students. In part (a) only a very few were unable to find the correct total from the table, making arithmetic slips. Nearly all were able to pick out at least one poor feature of the pictogram in part (b) with a significant majority managing two. Most common correct answers were the lack of a key and the different sizes of symbol used for the different age groups. Only a very few picked out as one of their two reasons the lack of a title or the unequal spacing (students should be aware that the symbols should be equally spaced.) A small number recognised that the implied scale of one symbol representing 250 children was poor. The most common incorrect comments were making reference to the unequal sized age groups. Some lost out on a second mark by making the same point in two different ways.

Question 2

Very few students failed to score in this question with most gaining a mark in part (a). Students seemed to be very aware that a hypothesis cannot take the form of a question. Whilst many pointed this out very concisely quite a number were less succinct, making reference to the specific wording, "do you think...", or to the question mark.

The idea of what constitutes a variable was less familiar to students although over half managed to score at least one mark in part (b). A lack of understanding was shown by lengthy descriptions of experimental design (eg "he must choose a variety of cities") often squeezed into the answer spaces, stating necessary equipment such as a thermometer, or by stating statistical terms such as "primary" or "discrete". The contextual idea of "height above sea level" was not clearly understood by a number of students with some referring to height of sea level ("sea level" alone was a common incorrect answer) and others referring to the need for a tape measure to find the height.

Fewer than half of students recognised the bivariate nature of the investigation and hence the need for a scatter diagram in answer to part (c). It was common for students to simply state the name of a diagram they knew (eg bar graph) with no apparent reason.

Question 3

This was a good question for most students with a significant majority gaining at least three marks and over half achieving full marks. Those who numbered the vertical scale generally did so correctly although a number missed this demand. A small number failed to include "percentage" in their label for the vertical axis and a very few simply labelled the axes x and y. Most students successfully added a bar for Spain correctly although often not too neatly.

If one mark was lost it was usually in part (d), often for simply listing the percentages for UK and USA or stating the difference was 32%. Questions often demand a comparison, as here, so simply listing is not sufficient. It may seem obvious to students that one listed value is higher than another but they must state this using comparative language of some sort. Another common wrong answer was to draw a comparison with Spain instead of USA.

Question 4

Another high scoring question for most. Students are generally quite good at completing two-way tables with very few making a slip in part (a). There were a small number however who were unable to interpret the table total for part (b), the most common error being to double, giving 90

Question 5

Students scored well in part (a) with 80% able to give a correct probability. Some simplified their answer to give $\frac{1}{5}$ although this was not required. A small number failed to score by giving their answer as a ratio. Part (b) was a little less successful although over half gave correct answers. One common error was giving 12, apparently finding the percentage rather than the number out of 50. Some failed to score by giving an answer of $\frac{6}{50}$ rather than 6

In part (c) students were split with just over half correctly opting for Edrik having the better estimate, most of these correctly referencing the larger sample in some way. For those scoring only one mark this was often due to poorly expressed or vague reasoning such as Edrik having a *wider* sample of people. Those who made the incorrect choice of Danni often suggested this was simpler due to the smaller sample.

Question 6

A small number of students failed to identify the correct value from the table in part (a); there was similar success identifying the correct age group in (c). Part (b) was more problematic although a good majority found the relevant figure of 86 from the table. Some had this as their answer; a variety of calculations were seen using the 86 although a small majority did correctly subtract from 100.

Students were less successful in part (d) where they had the commonly seen demand of identifying a trend. They need to know that stating figures at the start and at the end is not a description of trend. Whilst close to half were able to give an acceptable description of a rising trend it was clear that some had not read the question carefully, focusing on one gender or comparing age groups or genders. Some described the trend as 'up and down' which suggests they may have read the final row of the table as though it were a single sequence of values.

Question 7

Success on this question was very mixed. Many students remain unclear about the meaning of "population" with many stating "40" (population *size*) or simply "scouts". A minority knew "census" for part (b) albeit often poorly spelt, with incorrect answers including "survey", "questionnaire" or one from a selection of statistical terms. A further minority correctly recognised that the population was not large in giving their answer to part (c), (eg "there are only 40 of them") but common incorrect answers either gave a disadvantage of sampling or stated that the leader wanted information from all.

Typically in part (d) students focussed on standard problems with questionnaires rather than the context of using it with all 40 scouts, with "more time taken" being a popular answer. Common non-scoring answers were that the scouts

might all have different opinions or might not like the options offered. Part (e) was possibly the best answered part if the question was read carefully. (Some read less carefully and wrote a question on choice of activity.) The usual common error of missing or overlapping options was seen but there were many students scoring both marks here. Although not penalised some were perhaps not clear on the context with surprising time frames in their options up to multiple months for the summer camp.

Question 8

Interpreting the graph in parts (a) and (b) was where students were most successful in this question, although some had issues reading the scale in (b) with answers of 3.3 or 3.5 not being uncommon. The vast majority of students failed to realise in (c) that describing "the shape of the distribution" meant skewness had to be identified; "It goes up and then down" was often seen. Part (d) was a little more successful although only "about" one in three realised that for proportion to be shown they needed to select a pie chart.

Question 9

This question really challenged students. A correct answer of random sampling in part (a) was not uncommon but often few marks were gained in the attempts to describe taking a random sample from a large population in part (b). Some gave a definition of a random sample rather than how to take one. Marks gained in this QWC part were often for a reference to using random numbers or a description of using the corresponding function on a calculator. Other common sources of marks were for either numbering or listing the customers. There were a small number of students who gave the more technical answers of ignoring repeats or corresponding the random numbers to the customers to be selected for the sample.

There were many students who described "names in a hat" methods which are not appropriate for large populations. Other common errors were descriptions of systematic or convenience sampling, or not realising that it was the sampling method that needed describing rather than how to deal with the customers once chosen.

Question 10

Students generally scored well. It was usual to score at least one mark in part (a) for a description of the relationship but less common to score both. With two marks and the demand to "describe and interpret" students should realise there are two things to do. Most common was to omit the description of "positive correlation".

For part (b) a suitable line of best fit was most commonly drawn with only a minority not passing through the mean point as required. Pleasingly these were mostly ruled and sufficiently long. Similarly the two new points were usually

plotted with sufficient accuracy in part (c) although these were often generously rewarded if unlabelled; some students wrote A and B on the graph with no visibly plotted points. It was the reasoning in part (c) where students fell short when they failed to recognise the significance of the position of the points being above or below the line. A large number of students gained a mark for a correct choice of B, with a reason not related to the line as was required. Incorrect reasons commonly referred to closeness to the mean point. Some students failed to read the question correctly and considered all points rather than the two in the table.

Question 11

The unordered list caused problems for one third of students with often one of the minimum or maximum values being found incorrectly in part (a). Students should be encouraged to show working clearly as a subtraction for range with one of the values correct would have gained some credit. "Incorrect maths" was not uncommon here with the two values written in the wrong order for the subtraction; this did not score unless the student recovered with a correct answer. A small number attempted median or mean instead of range.

The frequency table and identification of modal class went well with full marks for about two thirds of students. The tally column was usually used appropriately in part (b) although there were some with slips losing one of the marks. There were a small number of students who did not know how to use the tallies or misunderstood the demand of the question as they completed a column with midpoints followed by another calculation such as cumulative frequency.

In part (d) students were provided with the necessary values to calculate the mean which the majority managed correctly. Unfortunately some ignored the given values and attempted to use the frequency table, usually without success. Incorrect calculations included dividing the stated frequency by number of class intervals. Far less successful were students' attempts to give an advantage of using mean in part (e). Commonly they just stated that it was accurate or easy to find, or that it gave the average, otherwise they described how it was calculated.

Question 12

This question was common with the higher Tier paper with most foundation students scoring between one and three of the five marks. The term "sampling frame" was not understood by most students so a mark was rarely scored in part (a). Incorrect attempts stated population or survey or described a sampling strategy. Part (b) was more accessible for many with the most common advantage given being obtaining results more quickly; "postal questionnaires might not be returned" was a common accepted converse comment. Fewer students scored well in part (c) with the most common correct answers describing pressure from the interviewer or lack of a telephone. Most gained

usually one of the two marks in part (c) for finding a total percentage of 80 but then failed to scale this up for the sample size of 1000

Question 13

Many foundation students scored quite well on this common question with higher Tier. In part (a) most gained at least one mark, usually for reading the chart in part (i), and over half gained full marks. The most common errors for part (a)(ii) were either stating 58 without subtracting the 20 for the lower bar, or for inaccurate reading of the scale. A small number read from the wrong composite bar chart. Some students wrote the subtraction incorrectly as 20 – 58 which did not score unless the correct answer was obtained.

In part (b) usually at least one mark was gained for an attempt with four sections in the correct order but poor accuracy or the bars stopping short of 100% meant no further marks were scored for some.

Students at this tier found part (c) more challenging with under a half gaining both marks. Where students did not score this was often with comments that referred to a single age group or for listing results without a comparison, or for comparison with the incorrect age group. For some only one comparison was made – students should recognise that two marks available indicates that two comments are required. Many referred to numbers rather than percentages although this was condoned in this case.

Question 14

Another question common with higher tier which was found challenging by foundation students. Over half were able to read the cumulative frequency graph correctly to score the first mark in part (a) but fewer could answer the second part where two values were required from the graph along with a subtraction. A single value read from the graph was common here as was subtractions in the wrong order. Many students drew lines on the graph to read from (which should be encouraged) although these were sometimes inaccurate (eg slipping a line). Some would gain more credit if they labelled values read from the graph and showed their subtraction.

Parts (b) and (c) proved too challenging for half of foundation students; there were a number left blank for at least one part. Some students who scored in part (b) commonly made correct reference to the fact that the survey was for a different population so may not be applicable. Some picked up on the small sample size. Incorrect answers often suggested it was the wrong graph to use or was inaccurate.

Part (c) was a commonly seen demand to compare distributions but not using this specific wording. Values also had to be found from the graph to enable the comparison. Only a small number of students gained full marks for correctly comparing medians and IQRs. More usual was gaining just one mark for a

comparison of medians. Often students did not find values of median or IQR for tablet owners to enable a comparison. Of those that did the values were often incorrect. A number of students were confused when reading the shape of the graph by making comments such as "more tablets are owned as people get older".

Question 15

Formal probability work usually challenges foundation tier students and this was the case here. There were a number of blank responses. The biased coin in part (a) was the first stumbling block with common incorrect answers being $\frac{1}{2}$, $\frac{0}{2}$ or a statement of likelihood. In part (b) students were less successful than usual as they had to draw their own probability tree rather than label a given one. Some managed trees of the correct shape and could gain full credit if using correctly an incorrect answer to (a). There was a mix of good and poor labelling of outcomes and probabilities. A sizable number of students did not know what to draw with answers including disjointed attempts at a tree or tables of various design. Finding an appropriate product in (c) was not common. When working was shown there was often an addition (which often included addition of denominators). Probability and tree diagrams clearly remains a topic for which foundation students do not have a good understanding.

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, United Kingdom