Examiners' Report

Principal Examiner Feedback

November 2020

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

GCSE (9 – 1) Statistics 1ST0

Principal Examiner feedback – Foundation Paper 1

Introduction

General Comments

This was the second entry of the new specification in GCSE Statistics. There were only a limited number of entries due to the extraordinary circumstances the year 2020 has presented. Once again, the questions in which candidates were asked to extract information from a table or a graph, or perform calculations were answered more successfully than the questions testing explanation and interpretation. The most challenging questions at this tier were the conditional probability [Question 11c] and the final extended response question [Question 14].

Centres should remind candidates of the need to read questions very carefully. A particular example of this was the Pie Chart question [Question 7] in which candidates were told explicitly in the second word of the first sentence that the Pie Chart was drawn accurately, and yet only a small minority of students actually measured the angle of the required sector.

Question 1

Part (a)

It was rare to see an incorrect response here, but even so, candidates should help themselves by writing totals by each bar in the chart and adding carefully. Though this part is worth one mark and so no marks are available for methods shown, it does help a candidate to write down $10 + 3$ to focus the mind.

Part (b)

The question states 'describe' and so we are looking for a description rather than a number, but a number of students wrote down 0 or 0% or $0/20$. We condoned these responses, but the correct word should be 'impossible'.

Part (c)

The majority of students were correctly able to identify that the population was the required statistical term.

Part (d)

An explanation was required in this part. Most candidates either wrote, that there were no teachers in Jon's sample, or conversely wrote that Jon only asked students, both responses being correct. Only a few students did not understand the question.

Question 2

Part (a)

In this part we not only required the total to be correct, but the tallying also needed to be correct using the standard notation [e.g. ⊬⊬⊢|||].

Once again, most candidates tackled this successfully although some omitted to find the total.

Part (b)

This was less successful, and it was clear that many candidates did not know how to find the mode from a set of data. Some candidates wrote down the frequency for Soya instead of the type of milk.

Part (c)

There are two elements to this question; firstly, determination of the mode in part (c), and then secondly to compare and comment on the mode in part (b) and part (c).

The answer we were looking for was that the mode on Tuesday is Almond, so it is not the same as the mode on Monday. We followed through the mode found in part (b), but even so, this was not well answered.

Question 3

The questions states: Discuss whether or not it is appropriate…

In other words, a comment is required whether or not the representation is appropriate, and then to give a reason for that decision.

Centres should encourage students to look at the total number of marks available for the question as this will give a hint on what is required. In this case there are three marks so, we awarded one mark for stating 'not appropriate' with some attempt at a reason and then one mark for each of two reasons.

The reasons in a question like this are fairly standard, but it helps candidates to imagine that the table with the raw data is not available and only the diagram is seen. In this case, there is no vertical scale, the diagram is in 3 dimensions, there are no labels and there is no title.

Question 4

Part (a)

Students were asked to extract a piece of information from the table and this was well answered.

Part (b)

This part also gained full marks in all but exceptional responses. It does help to double check values to the table by adding both columns **and** rows, although very few incorrect answers were seen.

Part (c)

This part was not answered well at all. Candidates need to decide whether the question is looking for a probability out of the total number, which is 94, or whether conditional probability is required.

In part (i) a description was asked for whereas in parts (ii) to (iv) the command words 'write down' or 'find' is telling the candidate that a number is required. We still saw many responses giving a description of 'likely/unlikely etc. If a description is required, the question will state 'describe' otherwise a numerical value must be given.

The word that indicates conditional probability is 'given' so the denominator in part (iv) cannot be the total of 94.

Question 5

Parts (a) to (c) require candidates to extract information from the table and (a) and (b) were answered very well. Part (c) however, was not answered well at all and it was obvious that candidates had not gone through and found the difference in each case.

Although there was more work in this part for one mark than in parts (a) and (b), with a calculator, these differences could be found very quickly and written into the table, yet no such notes were seen in the given table in responses, and very few correct responses were seen.

Part (d) was answered well.

Part (c) is another explanation/interpretation type question.

As before, a decision for which method is more appropriate is required together with an explanation. In quite a few cases, no decision was given meaning that no marks could be awarded.

Question 6

Part (a)

This part was well done overall. A few candidates merely repeated the question by stating; 'there is a relationship between a person's height and the person's salary' which cannot gain the mark.

Parts (b), (d) and (e)

This paper had a few questions testing definitions of standard statistical terms and candidates should know correct statistical words and their definitions. In this case only a few candidates were able to write down the word 'primary' in part (b), 'quantitative' in part (d) and 'quota' in part (e). We were generous on poor spelling, but nevertheless only a few correct responses in each part were seen.

Part (c)

Most candidates understood the requirements of the question and virtually all wrote a comment about confidentiality of people's earnings.

Part (f)

There were two demands in this part.

Firstly, to identify the calculation/column for which there was no error message for the age with an attempt at a supporting reason, and the second mark for identifying that height and salary needed to be cleaned.

Those candidates who understood the question did so very well, but those who did not understand the demand, attempted to perform calculations to find the mean for age, height or salary. We ignored any comment or calculations relating to the employee number.

Question 7

Part (a)

Most candidates were able to explain why fish made up the largest proportion, either referring to the area or explaining that white is the largest area.

Part (b)

The opening sentence states 'the accurately drawn pie chart...' but most candidates did not measure the angle which was $220°$. In fact many candidates assumed the angle was half a circle with half of a quarter, so in other words assumed the angle was $225°$.

We awarded one mark for those candidates achieving 33.75 million as a special case, which is the quantity arising from the assumption that the angle is $225°$.

The correct answer of 33 (million) was a rare sight in this group of candidates.

 Parts (c) and (d)

These parts were not at all well answered when even attempted, and virtually no candidate could interpret the upper quartile.

Part (e)

In this type of question where candidates are asked to comment on a conclusion, it is crucial to comment on whether or not the assertion is valid, and follow this up with an explanation. Quite a few candidates did not state 'not valid' but just gave a reason which could not score any marks.

Question 8

Part (a)

The question specifically states 'annual inflation rate' which meant that explanations referring to quarterly trends could not gain any marks. Centres should teach candidates to look at the mark allocation for questions, because as in this case, there are 2 marks. One mark is for describing the trend (i.e., trend is upwards after 2015) and one mark is for interpreting it (i.e, the annual inflation rate is increasing over the years). This style of question is not yet well understood by candidates.

Part (b)

This was done very well as expected with the points placed on the graph accurately in nearly every case.

Part (c)

A 'show' question means that every step of the calculation must be seen. In this case we needed to see $\dfrac{0.4+0.3+0.4+0.4}{4}$ or $\left(0.4+0.3+0.4+0.4\right)\div4$ but the following response which was seen occasionally $0.4+0.3+0.4+0.4\div4$ could not be awarded the mark as although the intention was clear, the calculation is mathematically incorrect.

Part (d)

This is another question requiring an explanation and it also requires a decision, (i.e., not suitable) to be part of the answer. Once again, the mark allocation is an indication to candidates. However, few candidates were able to answer this part at all.

Question 9

Part (a)

This is an extended response question as indicated by the allocation of 5 marks. Although the question did not mention the necessity for calculations, in order to test Cindy's, claim it was necessary to calculate the mean and the range; two calculations which very few candidates actually completed. As Cindy made two claims, then there needed to be two comments to either support or negate the claims.

There are at least three questions of this type in the Foundation examination and candidates need to be familiar in how to approach them.

Part (b)

Most candidates were able to state the correct reason here.

Question 10

Part (a)

This was another definition question and as mentioned earlier, candidates need to know the terminology of statistical words with their respective definitions.

Part (b)

The three marks allocated to this question means that three comments were required. Most students did not grasp the key point that 50 out of 467 students were required to be chosen using a random number method, and a common response was 'choose 50 students and give them each a number'. A number of responses referred to picking numbers out of a hat, meaning that the reference to a list of random numbers was required. No candidate managed to gain the mark for dealing with repeats/50 unique students required.

Question 11

Part (a)

This was mostly answered very well and only a few candidates wrote 5.2% instead of 4.1% to 9.3%.

Part (b)

Most candidates were able to identify and count the correct regions from the choropleth map.

Part (c)

Few candidates grasped the principle that a percentage change does not equate to an actual change if we do not know the quantities.

Part (d)

Very few candidates could name the Choropleth map used in this question.

Question 12

Parts (a) and (b)

The point about the estimate was lost on most candidates in this question, because most candidates read the new car registrations in Italy as 580, despite the bar being in between 560 and 580. This meant that they miscalculated the difference between the number of new car registrations in the UK and Italy, and were also unable to articulate the reason why the answer was an estimate.

Part (c)

The detail in the question was lost on some candidates, which emphasises the need to read questions very clearly. In a comparison we need to see a statement

that one is **more** or less such as; Italy has **more** alternative. Writing the quantities without this comparison gained no marks.

Part (d)

Virtually no candidate grasped the salient point of this question that long term predications cannot be made on the basis of a limited amount of data.

Question 13

Parts (a) and (b)

A pleasing number of candidates were able to complete the tree diagram correctly and find the probability in part (b) showing that they are well versed in this technique and calculation.

Part (c)

No candidate was able to answer this part correctly.

Question 14

Part (a)

At this stage of the paper is was pleasing to see how many candidates were able to articulate that it is appropriate that extreme values/outliers are removed prior to data processing.

Part (b)

This was the second extended question in the paper and required candidates to perform calculations and compare the summary statistics. In this case we had the data to find only the median and the interquartile range. The range, which was quoted by a number of candidates, was not accurate as the minimum and maximum values were not known for Paris and so could not be accepted.

Comparing just the lower quartiles or the upper quartiles do not score any marks.

As always a conclusion was required, for example;

Paris median $= 128$, London median $= 150$ so the London median is greater. Note that the use of the word median is essential here rather than the generic word average.

Summary

Based on their performance on this paper, candidates should:

- practice development of extended response questions, noting that calculations are frequently required to support any conclusions.
- use a protractor to measure angles in Pie Charts that are accurately drawn.
- develop the understanding that a large percentage change does not necessarily imply a large actual change.
- learn the names and definitions of statistical terminology carefully.
- give statistical interpretations in the context of questions.
- compare sets of data using any of the three measures of central tendency depending on the data or summary statistics given in the question, and either the range or interquartile range.