GCSE
STATISTICS

8382/2F: Paper 2 Foundation
Report on the Examination

8382
June 2022

Version: 1.0

Summary

Overall performance compared to last year
The performance was similar to June 2019. The state of handwriting was often very poor this year, making it really difficult to distinguish some numbers from each other and make out some words.

Areas/topics/skills where students excelled

- plotting missing points
- calculating a mean
- interpreting a pie chart to state the mode
- reading from a line of best fit
- commenting on trends shown by time series graphs

Areas/topics/skills where students struggled

- writing a hypothesis
- interpreting a pie chart to give the value for one sector
- plotting a double mean point
- choosing suitable calculations to perform, on an open question
- knowing when to use and implementing interpolation
- interpreting a cumulative frequency diagram and using the information to draw a box plot

Multiple choice questions

Which questions did students find most accessible

Questions 1 and 3 were the best answered multiple choice questions, showing a good understanding of exhaustive probability and the use of quartiles in an ordered set. Question 13 was the multiple-choice question that had the greatest number of non-attempts.

Which questions did students find most accessible

All multiple-choice questions were well attempted.

Of the non-multiple-choice questions, Question 6 was the question that the most candidates attempted.

Which questions did students find least accessible

Questions 11, 14 and 16 were the questions that more students found difficult.

Common misunderstandings

Question 5

A lot of students did not complete part c on Question 5 and that impacted their ability to fully answer part d.

Individual questions

Question 5

Parts a and b were each very well answered. The most common mistake in part a was to add to 2015 figures, to achieve a grand total for 2015. Loss of marks usually only occurred in part b when there was a slip in the subtraction seen.

Part c was not answered by quite a large proportion of the cohort and this meant they were unable to give a full answer in part d, unless they looked back at the original data in the table. We saw many comments that suggested that streaming only lasted for two years (the years 2013 and 2014 were already plotted on the graph). As in previous years, it was not uncommon to see a comparison between downloads and streaming for one particular year. It was pleasing to see that the vast majority of students who completed the graph used a dashed line to join the plots.

Question 6

In general, in part a, the plotting was done accurately. In part b, it was pleasing to see a high proportion of the cohort correctly calculating the mean (not much evidence of the median being attempted instead which can often happen). The usual error would have come from a slip in addition. Very few candidates purposely plotted the double mean point on the graph, so those that scored both marks in part c, often did so because their line of best fit was fortunate to pass through (23, 58).

For part d, if there was any positive line of best drawn on, and the candidate read off the value accurately, they could score the mark, whether or not it had been an acceptable line of best fit for part c. Incorrect answers were usually as a result of misreading the scale.

Most candidates realised that the new dog introduced for part e was unlikely to be a Vizla, even if their reasons were not clear enough. Most candidates tried to link either the mass of the new dog or the height of the new dog to the given data, but not both. It was extremely rare to see someone use the line of best fit to support their answer of No.

Question 7

Part a was extremely well answered with the usual error being to answer Facebook, as this was the second most popular social network platform. The most commonly used method in part bi was 0.8×68 and this was usually successful. Some students attempted to give reasons why Simran may/may not be correct with regard to real-world usage, rather than by performing appropriate calculations. Those who didn't score on part bii were trying to justify why people may/may not be on YouTube every day, instead of highlighting the issue of having no data to prove/disprove Simran's statement.

Question 8

It was common to see $84 \div 3$, instead of $84 \div 4$, although the 16 was almost always correctly placed, in part a. Occasionally, the 63 and the 21 were written in the wrong circles. It was most common to see the answer to part b given as a fraction, but some students converted to a decimal (incorrect conversions were not penalised). It was quite common to see 84 as the denominator, perhaps because the 16 wasn't seen as being part of the Venn diagram. The other common error

was to give the answer of $\frac{1}{3}$, perhaps because there are three distinct "spaces" on the Venn diagram.

The modified question paper version 36pt A3 for part a included an incorrect diagram. The senior examiner reviewed all of these modified scripts to ensure no candidate was disadvantaged.

Question 9

Students were comfortable answering this question, with some very pleasing scores. For part a, the common, incorrect answers were $\frac{1}{168}$ and $\frac{1}{9}$.

The students showed good knowledge of how a choropleth map should be presented. The error on the given choropleth map, most easily identified by the students, was the top, middle box being the wrong colour. It was great to see so many offering up "should be shaded lightest to darkest on this key" as one of the things that was wrong with the diagram.

Question 10

It was wonderful to see so many be able to extract the correct information from the pie chart for part a but a disappointing number in part b actually measured the relevant sector, mostly assuming it to be 90°. For those who did measure the angle, it was not uncommon to see the angle given as 80°, from reading the wrong side of the protractor. Two of the three marks were still available for correct method and answer from an incorrect angle.

Question 11

The students who understood how to complete the cumulative frequency column were usually headed straight for full marks across parts a and b. It was extremely rare to see a simple miscalculation in the cumulative frequency figures. Plots on the graph were usually accurate and at the end-points and generally joined with a smooth curve or straight lines (a fair mixture of each).

Part c did not go as well for the students, with few showing a good grasp of translating the information from the given cumulative frequency curve into a boxplot. Often the ends were plotted accurately but it was common to see the box drawn at 5, 15 and 25.

Question 12

In part (a), the students showed a good understanding of correlation not implying causation, in the given context. The answers were relating to both the rainfall and the sales of umbrellas.

In part (b) the most common answer was "Yes" with an explanation that Caro could draw on and then read off a line of best fit on the scatter diagram, not understanding that the 20mm would be outside the data set.

Question 14

Many students made a good start on identifying the issues with the stem and leaf diagram but did not go quite far enough to be awarded marks.

Part b was answered much better than part a, with almost half of the cohort scoring. The usual errors were to misplace the median (putting it between two 34's rather than landing on exactly one 34), to order only the leaves as one whole data set or to calculate the mean.

Question 15

It was clear from the answers to part a, that students had not read the information above the table to show the difference in figures between England and Wales. The most common answer for part a was 2018.

Part b was well answered with the majority of candidates correctly identifying the number of confirmed cases of mumps for England in 2011 and many of those then going on to correctly use the given formula. Mistakes seen were ×100 instead of ×1000 and not showing the interim value before the given final answer was written.

Question 16

The main reason that candidates did not score well on this question was that they only worked out the mean of the values and ignore the median and the mode. Some calculated the range, but that was no help to them, in this question. When calculations were seen, they were usually accurate. There was, as we often see, a lot of answers commenting on the individual best/worst times for the dogs, rather than average times. When commenting on the validity of the experiment for part b, most students commented about the possible lack of accuracy recording the times or said that there were too few dogs in the sample.

Question 17

The question in part a was answered in the typical way, with the majority of students posing a question or an opinion, rather than a hypothesis. The correct answers used the wording from the stem of the question, which was perfectly appropriate. Parts b and c were generally well answered with most of the students being able to identify at least one problem in part b and half of the cohort answering correctly in part c. The missing option for 70+ was the problem that most students identified. Those who did not score on these parts were usually answering that the questions had nothing to do with HS2.

On part d of the question, it was rare to get a full answer, from the few who knew how to take a random sample. Most of the students were detailing going to five stations and asking the first people they met. For those who made a start on a description of finding a random sample, it was not usual for them to include "without replacement" (or the equivalent) in their descriptions.

The most common, correct answer in part ei was "convenience" with "opportunity" the second most common. In eii and eiii, the most common answer referred to bias or a lack of bias. The students were not showing a good understanding of the questions asked, in this context. Part eiv saw a better set of answers, with students identifying that a broader range of answers would lead to better data.

In part f, few students were able to correctly name the diagram shown but they were able to do a reasonable job of reading the data from the dual bar chart and interpreting the time differences shown. There were many candidates however, who tried to answer the question without using values from the chart or rounded the values to fall out of an acceptable range.

On the final part of this statistical enquiry cycle question, most of the cohort correctly worked out the 44 minutes and those who went on to score marks for the 34.4% usually used the method

$$\frac{44}{128} \times 100$$

Mark Ranges and Award of Grades

Grade boundaries and cumulative percentage grades are available on the Results Statistics page of the AQA Website.