GCSE
STATISTICS

8382/2F Paper 2 Foundation
Report on the Examination

8382
June 2023

Version: 1.0

Further copies of this Report are available from aqa.org.uk

Summary

Overall performance compared to last year

The performance was similar to June 2019. There was a greater level of “not attempted” this year, compared to 2019 and last year.

Areas/topics/skills (delete as appropriate) where students excelled

- Simple probability
- Median of a small data set
- Completing a tally chart
- Reading from a table
- Describing a trend in data
- Completing a sample space diagram

Areas/topics/skills (delete as appropriate) where students struggled

- Explaining the advantage of a sample over a census
- Finding the number of items needed in one group for a particular stratification
- Calculating 4-point moving averages
- Finding the interquartile range from a stem-and-leaf diagram
- Knowing which values to calculate, in order to openly compare two sets of data
- Criticising an experiment

Multiple choice questions
Which questions did students find most accessible

Questions 1 and 3a were the best answered multiple choice questions.

Which questions did students find least accessible

Question 7b was the multiple choice question that had the greatest number of non-attempts.

Which questions did students find most accessible

Of the non-multiple choice questions, Questions 4b and 5a were the questions that the most students attempted.

Which questions did students find least accessible

Questions 11, 13bii and 13d were the questions that more students found difficult.

Common misunderstandings

Question 11

A lot of students found the difference between the values above the empty cells, meaning that they could not plot their values for part b, on the given graph.

Individual questions

Question 4

This was a well-attempted question. Roughly half of the cohort were able to express what “raw data” are and almost 90% of the students scored 2 or 3 marks on the tally. The most common, incorrect answers for part a were “data just collected” or “data collected by yourself”. Parts c and d fared equally well with almost half being able to tell us that the mode was used and three-quarters of students scoring on the question that judged Seb’s statement about the data. Many were choosing the correct number from the table, even if they weren’t quite able to put the data into comparable form.

Question 5

Part a was done very well, with around 90% of students able to identify the correct figures from the table, even if sometimes, their arithmetic went awry. Part b offered great responses where the general idea was indeed that the trend was decreasing. In part c, those who understood to find the difference between consecutive years rarely went astray. Some did answer for the whole table and some misunderstood which years their calculations lead them to. The most common approach for part d was to be working towards 103 and 125, with the next most popular method being to work towards 206. Many students found the 103 or 125 but were then unable to continue the process of proving whether or not that is more than one third. Part e was extremely well answered, in context.

Question 6

Common, incorrect answers to part a were “accurate”, “doesn’t cost money” and “reliable”. A good proportion of students were able to score at least one mark in part b, for identifying in context, reasons for this not being a good sample.

Question 7

Part a saw its share of misreading from the graph and this was the usual cause of not scoring the mark although many students were looking to answer a more complicated question. The majority of students understood what was expected of them. Students who didn’t answer correctly for part b usually chose option 1 or option 3. When drawing the frequency polygon in part c, the most common error was to plot at the ends of the intervals, rather than at the midpoints, despite the existing frequency polygon being there as an aid. It was pleasing to see so many students using the scale correctly. There was the expected misconception that the graphs showed the growths day-by-day rather than the heights or different plants and the expected misunderstanding that packet A had 5 plants each exactly 25cm tall etc.

Question 8

Part a was very well done with cell R6 being the one that went wrong, if any went wrong. Part b was well answered with some students picking up the SC1 for reading the instruction as “blue **and** 3 or less”. Part c was not well answered; those who appeared to begin well often ended up with $\frac{11}{24}$ instead of $\frac{9}{24}$.

Question 9

The reasons behind them choosing “secondary” were not always clear but the advantages/disadvantages of either type of data were less muddled. We saw a lot of “quicker” or “easier to obtain” for the advantage of secondary data and a lot of “takes a long time” or “costs a lot” for the disadvantage if they’d chosen primary data. In part b, knowing what to do after finding 709 was a struggle for many. Part c was quite well answered with students engaging well with the context. For those who didn’t score, it was mainly due to a reference to the number of passes, not the number of tests taken. Most students were able to detect the decreasing trend of the number of tests taken as the age increased for part d. When attempted, the calculations necessary in part e were accurate but sometimes the figures then drew the students’ attention away from the pass rate when making the final comment.

Question 10

Part a showed the usual muddling of explanatory and response variables. Part b scored disappointingly lowly with a huge mixture of answers from “eggs” to “hens” to “sample”. The reason required about stratification was answered contextually rather than using the figures in the table as a prompt, in the given context but very few were able to show the necessary calculation to prove that there ought to be 15 hens in the age group $2 \leqslant a < 4$.

Question 11

The 4-point moving averages were not done well, with the majority of students finding the midpoint of the two values above each empty cell, or the difference. This then impacted the plotting in part b, as differences wouldn’t fit onto the graph. Those who were able to plot their values, did so very well without much misreading of the scale.

Question 12

The formula for birth rate was used efficiently with only a small percentage of students using 100 instead of 1000. The comment in part b was not always answered well, with many students referring back to 2010 figures. Some that mentioned birth rates were also saying that higher birth rate meant greater number of births.

Question 13

Students that had understood the scenario usually answered well in part a, saying they'd need to take intelligence level into consideration. The median was found well enough by almost half of the cohort but the interquartile range was not something they found easy at all. The stem-and-leaf was put together well enough except for the key. The vast majority didn't understand how to write the key and for those that knew what was expected, they usually left off the percent signs. "compare statistically" continues to be a stumbling block for our students. There was little understanding that an average and a measure of spread would need to be calculated in order to compare. Most comparisons were done by using the maximum and minimum scores for each group. The criticism require in part e was also not easy for the students to express. It often came in the form of a suggested improvement for a future experiment.

Mark Ranges and Award of Grades

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