



GCSE **Statistics**

8382/1H

Report on the examination

Version: 1.0
June 2025

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

AQA retains the copyright on all its publications. However, registered centres for AQA are permitted to copy material from this booklet for their own internal use, with the following important exception: AQA cannot give permission to centres to photocopy any material that is acknowledged to a third party even for internal use within the centre.

General

The vast majority of questions were attempted by most students but with varying degrees of success.

Questions that were answered well included: drawing a box plot, calculating a 4-point moving average, drawing a stem and leaf diagram, completing, and interpreting a control chart and using capture/recapture to estimate a population.

Questions which were not answered well included: drawing and interpreting a cumulative frequency step polygon, identifying outliers by calculation, and determining the reliability of a normal distribution when given the skew and calculating basic probabilities.

A number of marks were lost through students not reading the questions correctly, but this was reflected in the final mark scheme when possible.

Students are reminded to show any working as many students failed to do this and as a result were awarded fewer marks.

Question 1

This was answered well with the majority of students referring to a fixed age range in the school or similar. Many incorrect responses simply stated that it would not be representative. Responses which did not mention ages were not awarded any marks.

Question 2

This was answered fairly well. The most common correct response was that there was no reason for the ants to be in that square. A number of incorrect responses mentioned that there was no food source or no nest. To be awarded a mark, responses needed to refer to both the food source and the nest.

Question 3

This was not particularly well-answered. Many of the incorrect answers compared either the total number of diesel cars to either the total number of cars or the total number of diesel vehicles. $\frac{72}{354}$ was a commonly seen incorrect response.

Question 4

This question was very answered well. The few incorrect responses usually referred to increasing the sample size.

Question 5

This question was common to the higher and foundation papers.

Part (a) was not answered well, although the majority of students were awarded at least one mark. The most common correct responses were that Q1 was always the lowest value, Q4 was always the highest value, and the income increased over time. Common incorrect responses referred to a single quarter or made a comment without context, eg the trend increases.

Part (b)(i) was answered very well. The most common incorrect response seen was to calculate the mean of the final four 4-point moving averages rather than the incomes for the final 4 quarters.

Part (b)(ii) was fairly well-answered. There were a number of other approaches taken which were covered in the mark scheme. The most common approach seen was to use the given 4-point moving average to calculate the income for Quarter 4 in 2019, obtaining a value of 88. Some students failed to state that Bob was incorrect which resulted in a mark being lost.

Part (c) was answered fairly well, and the majority of students were awarded at least 2 marks. Points were generally plotted correctly but the points were often joined together with individual lines rather than a single straight trend line. A number of students failed to plot their point from part (b)(i) and these responses could only score a maximum of 2 marks.

Question 6

This question was common to the higher and foundation papers.

Part (a) was well-answered, and almost all students were awarded at least one mark. Most students correctly identified the correct values of 5, 6 and 13 and wrote them in the table. There were few errors seen in plotting the students' values on the box plot. Students could still be awarded at least one mark if the values in their table were correctly used in their box plot. The quartiles were incorrectly stated as being 4 and 15 by a significant number of students, although often accompanied by the correct value for the median.

Part (b) was fairly well-answered with more than half of students being awarded at least two marks. A number of responses simply stated the values for the two medians but failed to interpret them in context. There were a number of responses that incorrectly used the ranges rather than the inter-quartile ranges. However, if the ranges were used and interpreted then a mark could be awarded.

Question 7

Part (a) was well-answered, and almost all students were awarded some marks. The common errors were to miss out a single data value or not arranging the data in order. These students were usually awarded 2 marks.

Part (b) was well-answered. The most common incorrect responses stated values for the medians or modal groups but did not interpret them.

Question 8

The majority of students were awarded at least two marks. There were some very good responses relating to the sunbeds, eg lack of information about frequency of use, and for using lotion, eg may not use it. Responses relating to people born in 1961 were weaker with the information being outdated being a common incorrect response.

Question 9

This question was not well-answered with over half of students failing to gain a single mark. A very common incorrect approach was to calculate the expected number of 6s from a single die (25), showing a lack of understanding of the situation described in the question. Many students were awarded 2 marks for $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$ and then tried compared it to $\frac{11}{150}$ but without changing these values into comparable forms. There were a number of different approaches seen and the mark scheme was amended to reward these methods.

Question 10

Part (a) was fairly well-answered, and the majority of students were awarded at least one mark, although few students scored 2 marks. It was not unusual for the first point to be correctly plotted but the next two to be incorrectly plotted at 10 and 15 on the cumulative frequency axis. These responses were awarded one mark.

Part (b) was not well-answered although the majority of students were awarded at least one mark. Students who correctly identified that 5 or more goals were scored in 15 games often went on to be awarded all three marks. It was not uncommon to read off values incorrectly from the diagram, however these students often calculated 20% of 64 correctly and so were awarded a single mark. It was common for students to incorrectly state that there was a total of 70 games.

Part (c) was not well-answered. A significant number of students were awarded a single mark for incorrectly doubling the correct probability of Harriet attending a single game. If $\frac{5}{64}$ was not explicitly seen, the special case of $\frac{10}{64}$ often resulted in the award of a single mark.

Question 11

Part (a) was not well-answered with less than half of all students awarded a mark. Many responses identified the correct two months but then failed to give a satisfactory reason, eg stating that the difference was 27.3 but then failing to state that this was the biggest difference. A number of responses simply stated the months without attributing to them the correct years. These responses were not awarded any marks.

Part (b) was fairly well-answered with the majority of students awarded some marks. There were many different approaches taken when answering this question and these were all covered by the mark scheme. Calculating the percentage change was the most common successful method. The most common incorrect method was to find the difference between the two values for Somerset and then incorrectly state that this was 3.3%

Question 12

Part (a) was less successfully answered with few students awarded marks. The most common response was to state that the diagram showed age and gender which was not awarded the mark. Correct responses often referred to the higher percentages of males in a specific age range.

Part (b) was fairly well-answered with the majority of students awarded at least one mark. Most students correctly calculated that the percentage of males in the specified age range was 28% but then calculated 28% of the 4200 rather than using the fact that 4200 was 28% of the total population. A number of students used a different age range to the one stated which was not awarded any marks.

Part (c) proved challenging to many. A common misconception was to state that the value of the 10th percentile had to be less than 19. The most common correct response seen was $60 - 19 = 41$, realising that the 10th percentile could be as high as 19 and the 90th percentile had to be greater than 59, meaning 60 was the lowest possible value in this context.

Question 13

This question was well-answered. Most students correctly set up an equation, but a number could not correctly rearrange their equation and were restricted to having a single mark awarded.

Question 14

Part (a) was well-answered with most students correctly stating an answer in the given range.

Part (b)(i) was fairly well-answered. The majority of correct responses either calculated the change in the frequencies or stated that Country A had the steeper line. The most common incorrect response was to state that the line for Country A was lower or similar.

Part (b)(ii) was fairly well-answered. Many students realised that they needed to read off where the two lines crossed but then incorrectly chose to read from the cumulative frequency axis.

Question 15

Part (a) was not well-answered. A significant number of students failed to make any attempt at this question. Many students gave reasons that lacked context, eg there are only two outcomes rather than each pack either has a rare card or it does not. If students gave two reasons without context, they were awarded a single mark.

Part (b) was not well-answered. Most successful responses started by calculating the probability of no rare cards and subtracting from 1. A number of students stated the probability of getting a rare card was 0.5 but the most common incorrect answer was $5 \times 15\% = 75\%$.

Question 16

Part (a)(i) was very well-answered.

Part (a)(ii) was well-answered. The lower action limit had more errors than the upper warning limit, with the line for the lower action limit drawn at either 25 or 26.

Part (b) was very well-answered.

Part (c) was fairly well-answered with the majority of students awarded at least one mark. Many students correctly stated that the machine should be stopped or reset but then failed to give a reason, ie the value was above the upper action limit. These responses were awarded one mark. The most common incorrect response stated that another sample should be taken

Question 17

This question was fairly well-answered with the majority of students awarded the mark.

Question 18

Part (a) was fairly well-answered. Most of the correct responses referred to it being easier to use the data rather than collecting it yourself. The common incorrect response was that the data would be reliable because it had been collected by a doctor.

Part (b)(i) was not well-answered. Many students correctly identified the correct row to remove but failed to include reference to the mass being omitted in their reason.

Part (b)(ii) was well-answered. Most students correctly identified row 48 but a significant number incorrectly stated that it should be removed because the mass was an outlier.

Part (c) was not well-answered but the majority of students scored at least two marks for correctly substituting into the given formula and correctly calculating the standard deviation. It was far less common for students to calculate the mean. Students who correctly calculated the standard deviation and the mean often struggled to recall the formulae for determining if there were any outliers. It was not unusual to see responses involving 1.5 or 2 standard deviations rather than 3.

Part (d)(i) was answered less successfully. Incorrect responses usually failed to reference the curves or made an incorrect comment about the curves, eg the Gentoo's peak was higher (implying the height of the curve rather than the mass of the penguins).

Part (d)(ii) was fairly well-answered. The most commonly seen correct response was to state that the curve for the Gentoo penguins was wider. Many responses gave values for the ranges of each type of penguin but unless this was linked to the width of the curves, these responses were not awarded a mark.

Part (e) was fairly well-answered. The commonly stated correct reasons were that the value was furthest away from 0 or the mean or that 2.87 was bigger than 2.42. The most common incorrect reasons stated that the value was negative or that the mean had a value of 1.

Part (f) was not well-answered. Many students simply listed the reliabilities of each type of penguin based on their skew but did not state that it was unreliable for the Gentoo due to the large value for the skew. Many students incorrectly stated that the Gentoo was the most reliable due to the high value for its skew.

Mark Ranges and Award of Grades

Grade boundaries and cumulative percentage grades are available on the [Results Statistics](#) page of the AQA Website.