

STATISTICS

Paper 4040/12 Paper 1
--

Key messages

Where different approaches are possible in calculating statistical measures the most economical method should be used.

In a given practical situation candidates should consider if an answer they have obtained is a reasonable one for that situation.

Learned statements such as 'it is not affected by addition and subtraction' have very limited meaning on their own. They should be expanded to relate to the specific variables in the specific context of the question.

It is often in answers to the writing parts of questions that the extent of statistical understanding is demonstrated.

General comments

Good quality work was again seen on those questions involving routine numerical calculations. This was particularly the case with finding crude and standardised rates (see **Question 5** below), calculating the mean and standard deviation of a grouped frequency distribution (see **Question 8** below), and finding the equation of a line of best fit to given bivariate data (see **Question 10** below).

Performance on the shorter questions, other than that on crude and standardised rates, was more mixed. It was most limited on the statistical measures on small data sets (see **Question 2** below) and on the two-way table (see **Question 6** below). There were few fully correct answers to the probability question (see **Question 4** below).

Candidates again had difficulty with several of the writing and 'explain' questions. It sometimes seems as though short statements have been learned, which might not be fully understood, but which are reproduced in the hope that they might be appropriate in answering these types of questions. These statements are often so general as to convey little meaning in relation to the specific context of the question (see **Question 8(e)** and **9(e)** below).

As has been emphasised repeatedly in these reports, Statistics is a subject with practical context, used to analyse real-life situations. It is not just about manipulating abstract numbers; the numbers relate to variables in the real world. Candidates should look at their answers and try to think whether or not they are reasonable in the given situation. There was a particular instance in this paper where totally unrealistic answers were sometimes seen. Had the candidate appreciated this they would have been able to look at their working again, find what was wrong and correct it (see **Questions 9(d)(i)** and **(ii)** below).

Comments on specific questions

Question 1

A fair number of candidates made fully correct choices. The most common error was to choose 'survey' instead of 'census' in **part (c)**.

Question 2

Almost all candidates had fully correct responses to **parts (a) and (b)**. Very few candidates had fully correct responses to **parts (c) and (d)**. What usually prevented marks being awarded here was that the total number of attendees at these ten committee meetings was never found; the numbers 15, 14, 15, 12 and so on, were rarely seen. Instead, the two sets of data were incorrectly combined to form a set of twenty, even though ten numbers in the resulting list were for the teachers and ten for the students. Completely erroneous calculations were then carried out on the twenty numbers.

Question 3

To be able to interpret a Venn diagram successfully it is essential to understand what each of the seven regions of the diagram represents. A fair number of candidates showed such understanding and obtained the first three marks. Completing the diagram in **part (b)** proved more challenging, but few responses earned no marks. Where only one mark was awarded it was often for the change from the original diagram following from the information given in the third bullet point.

Question 4

There was a modest number of fully correct answers to the first three parts. Errors were usually in **parts (a) and (c)**. In **part (a)**, 'contains 4 chords' seems to have been interpreted sometimes as 'contains at least 4 chords'. In **part (c)** it was commonly understood that the denominator was now 9 instead of 15, but the numerator was incorrect.

In **part (d)**, although the question did not say so explicitly, the candidate had to interpret the choices being made in this situation as being choices made without replacement. Many did not do this, presenting the product of a pair of fractions with both denominators of 15. Some who did understand it was a without replacement situation did not consider the possibility of not F/F as well as F/not F. A few needlessly long solutions were seen, with mixed success, considering the six possible combinations of the groups of songs with the chord and those without.

Question 5

This question was very well done with many fully correct answers being seen to the first three parts. Candidates showed good understanding of how to calculate crude and standardised rates. Much less understanding was shown in **part (d)** of what it is about a population that makes these rates equal. Many candidates seemed to think it was something to do with the relative death rates amongst the different age categories of the population. Very few referred to the structure of the population.

Question 6

The first three parts of the question were answered reasonably well. In particular, some good explanations were seen in **part (a)**, pointing out the impossibility of someone visiting the café without visiting the shopping centre. Answers to **part (c)** were more mixed. Many correct answers were seen, but many also where the frequencies only in the table were added. Thus it was common to see answers coming from $9 + 10 + 5$ instead of from $9 \times 1 + 10 \times 2 + 5 \times 3$.

The last two parts were the least well done. In **part (d)** a common incorrect answer was that found from the row total of one visit. In **part (e)** very few candidates referred to the sampling frame which it is necessary to have in order to select a random sample.

Question 7

At the start of the question the footwear value for 2020 was given as '245 million dollars', and at the end of each of the first four answer lines were the words 'million dollars'. So it was totally unnecessary for candidates to write out, and show working throughout with, 245 000 000 and similar numbers with several zeros.

Almost all candidates had fully correct answers to **part (a)(i)**. Good answers to **part (a)(ii)** showed clear steps in the working, the first being to find the total footwear value for 2024, then this value being used with the angle on the chart for Asia to complete the solution. Answers which used a total value of 245 million

dollars also for the year 2024 could not be awarded any marks. The 2024 chart is clearly smaller than the 2020 chart, and, as these are comparative pie charts, it must represent a smaller total.

Most candidates were able to use the percentage bar charts properly in **parts (b) and (c)**, though mistakes were sometimes made when the wrong chart of the pair was used for the question asked. A fairly common occurrence in **part (c)** was to use correct percentages taken from the charts with interchanged correct totals for the two years. **Part (d)** involved use of both the 2020 pie chart and the 2020 percentage bar chart, and there were few fully correct answers. But many candidates were able to obtain partial credit for obtaining either the numerator or the denominator of the expression needed to find the required percentage.

Question 8

In **part (a)** almost all candidates showed competence in computing the mean and standard deviation of a grouped frequency distribution. But much variation was seen in the economy of the calculation methods adopted. By far the best were those which used the blank part of the data table to create columns for x , fx , x^2 and fx^2 , with the totals of fx and fx^2 then calculated and inserted into the relevant formulae. The least economical were those finding the standard deviation from the version of the formula using $\sum f(x - \bar{x})^2$. With this approach much of the answer space was taken up with working comprising many numbers with many decimal places. Such solutions would have taken much more time to write.

For the histogram in **part (b)**, answers fell into two rough categories: those taking account of the different widths of the classes, and those simply drawing the rectangles at the heights of the table frequencies. The actual amount of working required in **parts (c) and (d)** was quite small, yet many candidates filled the answer space with spurious working, sometimes involving class widths, and gained little credit. The difficulty seems to have been that many did not realise that the key number required in both parts, the total hours for which the charging point was used, had already been found in **part (a)**: it was $\sum fx$. The best answers seen to **part (e)** were those which demonstrated thought about the practical context of the problem: answers such as 'we do not know the charging time of each vehicle'. Much less satisfactory were brief, vague responses such as 'it is grouped data'.

Question 9

Many answers to **part (b)** earned only one or two of the three available marks. The question had stressed that a cumulative frequency *polygon* had to be drawn, so the plotted points had to be joined with straight lines, not the smooth curve which was often seen. A more serious error was made by those candidates who plotted points at the mid-class values instead of at the upper class boundaries. This necessarily affected the accuracy of later answers.

In **part (c)** candidates displayed good knowledge of how to use the graph to find these statistical measures. Many explained their work clearly by drawing lines on their graphs, and this enabled examiners to award marks for method even when the graph had not been drawn correctly and the numerical answer was wrong. It is good practice to do this throughout any question involving a cumulative frequency graph.

In **part (d)** many candidates do not seem to have read the question properly, because they assumed the given temperature rise occurred every year, not every ten years as stated. Few fully correct answers were seen here. The two questions were such that a candidate could judge whether or not the answers they found were at least reasonable. Yet totally unreasonable answers were commonly seen. Thus for example in **part (d)(i)** the increase in sea temperatures might be said by some to produce a lower median than before, or even a median of several hundred degrees. In **part (d)(ii)** the number of days in the year 2084 for this particular event to occur was occasionally said to be several thousand.

In **part (e)** explanations as to why the interquartile range would not change were often inadequate. Short statements of a very general nature, not related to the context, were often given, such as 'it is not affected by addition and subtraction'. There is no 'subtraction' (= decrease in sea temperature here), and the 'addition' is unexplained in relation to the change in the value of the quartiles as the sea temperature increases. In **part (f)** very few candidates were able to make the very simple observation that the plotted points had been joined (or should have been joined) by straight lines and not a curve.

Question 10

Very good work was seen in the first three parts of this question, with many candidates finding the equation of the line of best fit accurately. A few used points other than the averages to do this; whilst the method is correct it usually results in a less accurate answer. Some equations were seen which, with pause for thought, the candidate might have realised had to be wrong. The plotted points clearly show that the gradient is negative, and the y-axis intercept is somewhere near to +0.5.

Few candidates obtained full marks on the two parts of **part (d)**. Many started correctly by substituting the distances into the line of best fit. But then in **part (d)(i)**, for example, the resulting value of y was commonly multiplied by 2 and 20, in no particular order, or by 40, and an answer of 12.2 given. Only partial credit could be given for this. Often it seems not to have been understood that the value of y coming from the equation was a probability. This had first to be multiplied by 20 and rounded to the nearest integer to find an estimate for the number of successful shots. The estimated number of shots could then be multiplied by 2 to find the estimated number of points, an integer.

In **part (e)** a fair number of candidates understood that this was about the distinction between interpolation and extrapolation. But a few chose the 5.5 m distance because they said there was a better chance of success from the shorter distance.

STATISTICS

Paper 4040/13 Paper 1
--

Key messages

Where different approaches are possible in calculating statistical measures the most economical method should be used.

In a given practical situation candidates should consider if an answer they have obtained is a reasonable one for that situation.

Learned statements such as 'it is not affected by addition and subtraction' have very limited meaning on their own. They should be expanded to relate to the specific variables in the specific context of the question.

It is often in answers to the writing parts of questions that the extent of statistical understanding is demonstrated.

General comments

Good quality work was again seen on those questions involving routine numerical calculations. This was particularly the case with finding crude and standardised rates (see **Question 5** below), calculating the mean and standard deviation of a grouped frequency distribution (see **Question 8** below), and finding the equation of a line of best fit to given bivariate data (see **Question 10** below).

Performance on the shorter questions, other than that on crude and standardised rates, was more mixed. It was most limited on the statistical measures on small data sets (see **Question 2** below) and on the two-way table (see **Question 6** below). There were few fully correct answers to the probability question (see **Question 4** below).

Candidates again had difficulty with several of the writing and 'explain' questions. It sometimes seems as though short statements have been learned, which might not be fully understood, but which are reproduced in the hope that they might be appropriate in answering these types of questions. These statements are often so general as to convey little meaning in relation to the specific context of the question (see **Question 8(e)** and **9(e)** below).

As has been emphasised repeatedly in these reports, Statistics is a subject with practical context, used to analyse real-life situations. It is not just about manipulating abstract numbers; the numbers relate to variables in the real world. Candidates should look at their answers and try to think whether or not they are reasonable in the given situation. There was a particular instance in this paper where totally unrealistic answers were sometimes seen. Had the candidate appreciated this they would have been able to look at their working again, find what was wrong and correct it (see **Questions 9(d)(i)** and **(ii)** below).

Comments on specific questions

Question 1

A fair number of candidates made fully correct choices. The most common error was to choose 'survey' instead of 'census' in **part (c)**.

Question 2

Almost all candidates had fully correct responses to **parts (a) and (b)**. Very few candidates had fully correct responses to **parts (c) and (d)**. What usually prevented marks being awarded here was that the total number of attendees at these ten committee meetings was never found; the numbers 15, 14, 15, 12 and so on, were rarely seen. Instead, the two sets of data were incorrectly combined to form a set of twenty, even though ten numbers in the resulting list were for the teachers and ten for the students. Completely erroneous calculations were then carried out on the twenty numbers.

Question 3

To be able to interpret a Venn diagram successfully it is essential to understand what each of the seven regions of the diagram represents. A fair number of candidates showed such understanding and obtained the first three marks. Completing the diagram in **part (b)** proved more challenging, but few responses earned no marks. Where only one mark was awarded it was often for the change from the original diagram following from the information given in the third bullet point.

Question 4

There was a modest number of fully correct answers to the first three parts. Errors were usually in **parts (a) and (c)**. In **part (a)**, 'contains 4 chords' seems to have been interpreted sometimes as 'contains at least 4 chords'. In **part (c)** it was commonly understood that the denominator was now 9 instead of 15, but the numerator was incorrect.

In **part (d)**, although the question did not say so explicitly, the candidate had to interpret the choices being made in this situation as being choices made without replacement. Many did not do this, presenting the product of a pair of fractions with both denominators of 15. Some who did understand it was a without replacement situation did not consider the possibility of not F/F as well as F/not F. A few needlessly long solutions were seen, with mixed success, considering the six possible combinations of the groups of songs with the chord and those without.

Question 5

This question was very well done with many fully correct answers being seen to the first three parts. Candidates showed good understanding of how to calculate crude and standardised rates. Much less understanding was shown in **part (d)** of what it is about a population that makes these rates equal. Many candidates seemed to think it was something to do with the relative death rates amongst the different age categories of the population. Very few referred to the structure of the population.

Question 6

The first three parts of the question were answered reasonably well. In particular, some good explanations were seen in **part (a)**, pointing out the impossibility of someone visiting the café without visiting the shopping centre. Answers to **part (c)** were more mixed. Many correct answers were seen, but many also where the frequencies only in the table were added. Thus it was common to see answers coming from $9 + 10 + 5$ instead of from $9 \times 1 + 10 \times 2 + 5 \times 3$.

The last two parts were the least well done. In **part (d)** a common incorrect answer was that found from the row total of one visit. In **part (e)** very few candidates referred to the sampling frame which it is necessary to have in order to select a random sample.

Question 7

At the start of the question the footwear value for 2020 was given as '245 million dollars', and at the end of each of the first four answer lines were the words 'million dollars'. So it was totally unnecessary for candidates to write out, and show working throughout with, 245 000 000 and similar numbers with several zeros.

Almost all candidates had fully correct answers to **part (a)(i)**. Good answers to **part (a)(ii)** showed clear steps in the working, the first being to find the total footwear value for 2024, then this value being used with the angle on the chart for Asia to complete the solution. Answers which used a total value of 245 million

dollars also for the year 2024 could not be awarded any marks. The 2024 chart is clearly smaller than the 2020 chart, and, as these are comparative pie charts, it must represent a smaller total.

Most candidates were able to use the percentage bar charts properly in **parts (b) and (c)**, though mistakes were sometimes made when the wrong chart of the pair was used for the question asked. A fairly common occurrence in **part (c)** was to use correct percentages taken from the charts with interchanged correct totals for the two years. **Part (d)** involved use of both the 2020 pie chart and the 2020 percentage bar chart, and there were few fully correct answers. But many candidates were able to obtain partial credit for obtaining either the numerator or the denominator of the expression needed to find the required percentage.

Question 8

In **part (a)** almost all candidates showed competence in computing the mean and standard deviation of a grouped frequency distribution. But much variation was seen in the economy of the calculation methods adopted. By far the best were those which used the blank part of the data table to create columns for x , fx , x^2 and fx^2 , with the totals of fx and fx^2 then calculated and inserted into the relevant formulae. The least economical were those finding the standard deviation from the version of the formula using $\sum f(x - \bar{x})^2$. With this approach much of the answer space was taken up with working comprising many numbers with many decimal places. Such solutions would have taken much more time to write.

For the histogram in **part (b)**, answers fell into two rough categories: those taking account of the different widths of the classes, and those simply drawing the rectangles at the heights of the table frequencies. The actual amount of working required in **parts (c) and (d)** was quite small, yet many candidates filled the answer space with spurious working, sometimes involving class widths, and gained little credit. The difficulty seems to have been that many did not realise that the key number required in both parts, the total hours for which the charging point was used, had already been found in **part (a)**: it was $\sum fx$. The best answers seen to **part (e)** were those which demonstrated thought about the practical context of the problem: answers such as 'we do not know the charging time of each vehicle'. Much less satisfactory were brief, vague responses such as 'it is grouped data'.

Question 9

Many answers to **part (b)** earned only one or two of the three available marks. The question had stressed that a cumulative frequency *polygon* had to be drawn, so the plotted points had to be joined with straight lines, not the smooth curve which was often seen. A more serious error was made by those candidates who plotted points at the mid-class values instead of at the upper class boundaries. This necessarily affected the accuracy of later answers.

In **part (c)** candidates displayed good knowledge of how to use the graph to find these statistical measures. Many explained their work clearly by drawing lines on their graphs, and this enabled examiners to award marks for method even when the graph had not been drawn correctly and the numerical answer was wrong. It is good practice to do this throughout any question involving a cumulative frequency graph.

In **part (d)** many candidates do not seem to have read the question properly, because they assumed the given temperature rise occurred every year, not every ten years as stated. Few fully correct answers were seen here. The two questions were such that a candidate could judge whether or not the answers they found were at least reasonable. Yet totally unreasonable answers were commonly seen. Thus for example in **part (d)(i)** the increase in sea temperatures might be said by some to produce a lower median than before, or even a median of several hundred degrees. In **part (d)(ii)** the number of days in the year 2084 for this particular event to occur was occasionally said to be several thousand.

In **part (e)** explanations as to why the interquartile range would not change were often inadequate. Short statements of a very general nature, not related to the context, were often given, such as 'it is not affected by addition and subtraction'. There is no 'subtraction' (= decrease in sea temperature here), and the 'addition' is unexplained in relation to the change in the value of the quartiles as the sea temperature increases. In **part (f)** very few candidates were able to make the very simple observation that the plotted points had been joined (or should have been joined) by straight lines and not a curve.

Question 10

Very good work was seen in the first three parts of this question, with many candidates finding the equation of the line of best fit accurately. A few used points other than the averages to do this; whilst the method is correct it usually results in a less accurate answer. Some equations were seen which, with pause for thought, the candidate might have realised had to be wrong. The plotted points clearly show that the gradient is negative, and the y-axis intercept is somewhere near to +0.5.

Few candidates obtained full marks on the two parts of **part (d)**. Many started correctly by substituting the distances into the line of best fit. But then in **part (d)(i)**, for example, the resulting value of y was commonly multiplied by 2 and 20, in no particular order, or by 40, and an answer of 12.2 given. Only partial credit could be given for this. Often it seems not to have been understood that the value of y coming from the equation was a probability. This had first to be multiplied by 20 and rounded to the nearest integer to find an estimate for the number of successful shots. The estimated number of shots could then be multiplied by 2 to find the estimated number of points, an integer.

In **part (e)** a fair number of candidates understood that this was about the distinction between interpolation and extrapolation. But a few chose the 5.5 m distance because they said there was a better chance of success from the shorter distance.

STATISTICS

Paper 4040/22 Paper 2
--

Key messages

To be successful in this examination the context of a question should be carefully considered, and interpretation of diagrams, graphs and calculations should always be made with this context in mind. Final answers should also be checked to ensure that they seem sensible in the given context.

Candidates should read questions carefully and provide clearly set out solutions, particularly in questions that begin 'Show ...' and in multistage problems.

General comments

It is important that interpretation of results is made with the context in mind. In **Question 8dii**, for example, it was important to give examples specifically in the context of the hotel and avoid generalised answers that could apply to any context. An example of an appropriate answer was 'the amount of food bought might have changed'. A generalised answer that could apply to any context, such as 'the quantity of items changed', should be avoided.

Checking a final answer in the context of the problem is an important stage in the working. A candidate who had incorrectly divided by 100 rather than by 26, in **Question 8c**, for example, would have reached an answer of the wrong sort of magnitude, that, when considered as representing the overall increase in costs for the hotel, would not have been compatible with the calculated price relatives. Similarly, in **Question 8di**, candidates who divided rather than multiplied by a weighted aggregate cost index that was greater than 100, say, would have reached an estimate for expenditure in 2023 that was lower than the figure for 2022. Considering whether their answer was sensible or not, in the given context, could have alerted them to their error.

Clearly set out solutions were important in many of the questions. **Question 6bii** asked candidates to 'Show ...', for 3 marks. The three required stages in the working, namely, showing how many diesel cars should be in the sample, showing how many diesel cars were in the sample and drawing the appropriate conclusion should be clearly set out.

Some of the multistage problems, such as **Questions 7d** and **10c**, often had clearly set out solutions where candidates who had made errors could receive partial credit because their method was clear. For many candidates, more clearly set out working in the form of probability tree diagrams might have been helpful in **Question 9bi** and **Question 9ci**.

Comments on specific questions

Question 1

Most candidates made a reasonable attempt at this question, although fully correct solutions were rarely seen. Most candidates were able to correctly find the required measures for the most straightforward of the continuous variables, namely the length of time. The other continuous variable, age, had the complication of being in completed years, and some candidates either ignored this aspect or dealt with it incorrectly. Answers for the discrete variable, the number of visitors, were most likely to be incorrect.

Question 2

- (a) (i) Most candidates were able to interpret the sectional bar chart correctly. The small number of errors in this question appeared to come from candidates incorrectly reading values from the vertical scale, but this was very rare.
- (ii) Fully correct solutions to this part were seen in most cases, with just a few candidates finding either the number, rather than the percentage, or finding the percentage who said Internet, rather than Newspapers. Reading the question carefully might have prevented these errors.
- (b) (i) Able candidates were able to give an answer equivalent to 'it appears to be true' because '25 per cent of adults and 31.25 per cent of children use TV and Radio' or because 'the proportion (or percentage) of adults using TV and Radio is smaller than the proportion of children using TV and Radio'. Very few candidates gave a correct comment about the sample size or the reliability of the sample. Many candidates instead suggested that Neo was incorrect based on the numerical values of 10 children and 12 adults, without taking account of the different population sizes.
- (ii) Candidates who had correctly considered percentages in **part (b)(i)** did not always suggest a type of chart that would display percentages, with a dual bar chart being a common incorrect response. Charts that would have allowed Leo to make a comment about likelihood more conveniently would have been, for example, comparative pie charts or a percentage sectional bar chart. Other incorrect responses seen included incomplete answers such as 'percentage bar chart'.

Question 3

- (a) Many candidates demonstrated a good understanding of linear interpolation, with a variety of correct methods leading to a correct median. Those that made an error had generally found the correct class interval for the median journey length.
- (b) It was very common for candidates to assume that the interquartile range would be impacted by the addition of a constant to each piece of data as well as by the multiplication of each piece of data by a constant, with an incorrect answer of \$12.20 being very common. A measure of spread only being impacted by the multiplication by a constant was understood by only a very few candidates.
- (c) Using linear interpolation to find a number of journeys from a given distance proved more difficult than finding the median distance. The most common incorrect answer seen was 95, the number of journeys shorter than, rather than longer than, 18 km. Rereading the question might have prevented this error.

Question 4

- (a) Candidates tended to be more successful in listing the independent events than the mutually exclusive events. The most common error seen in the list of independent events was either one or two of the four pairs omitted, or the inclusion of A and B, the pair of mutually exclusive events, in the wrong list.
- (b) Most candidates were able to score at least some of the marks in this question, with $P(B \text{ and } C)$ being more often found correctly. The most common error was for the two events to be treated as if they were mutually exclusive when it came to finding $P(B \text{ or } C)$, with the subtraction of $P(B \text{ and } C)$ being omitted. A few candidates confused the words 'or' and 'and', incorrectly finding $P(B \cap C)$ for $P(B \text{ or } C)$ and finding $P(B \cup C)$ for $P(B \text{ and } C)$.
- (c) The most common incorrect answer was 4 rather than 5. For the red bag it is known that $P(\text{odd}) = 0.6$ or $3/5$, so there must be at least 5 discs in that bag.

Question 5

- (a) This question was answered well by most candidates. Those who did not score full marks often scored a method mark for having a correct expression for the number of standard deviations below the mean.

- (b) There were many fully correct solutions seen to this part. The most common errors were for a numerical value to be used for one occurrence of the unknown in the equation $\frac{x-23}{6} = \frac{x-30}{9}$, or for different letters to be used, x and y , say, for the two occurrences of the unknown. There were sometimes other expressions/methods, equivalent to the one given above, that were seen.
- (c) This proved to be the most difficult of the three parts to this question, although many fully correct solutions were seen. Some candidates formed a correct initial equation but were unable to solve it correctly. Others tried to find the standard deviation by using the formula $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$, an incorrect method for this question.

Question 6

- (a) In this question it was necessary to adapt the given two-digit random number table to produce equally likely numbers in the range 000 to 359, by dividing the table into 3-digit blocks. Candidates found this question challenging, with most producing a list of two-digit rather than three-digit numbers.
- (b)(i) Many candidates selected a stratified sample in this part, although some selected a sample stratified by both fuel type and gearbox, rather than by gearbox alone. A small number selected a simple random sample, without stratifying.
- (ii) Having selected their sample, many candidates correctly identified which of their sample were diesel cars, usually by listing them. Most also showed the necessary calculation for finding the number that would be in a representative sample, although some candidates did one or other of these without clearly showing both. It is important in a question where candidates are asked to 'show' a particular result, that all the necessary working is clearly laid out. Those that had a correct sample in **part (b)(i)** often provided all the necessary working and drew the correct conclusion from that working.

Question 7

- (a) It was pleasing to see many fully correct descriptions of the key. The most common errors were to only describe the part of the key on the right, so just 17 per cent in 2020, or to incorrectly define the value on the left as 1 per cent in 1960 rather than 10 per cent in 1960.
- (b)(i) Most candidates used the key correctly and found the required medians and ranges. The most common errors came from either slips in locating one or both median values or for errors in finding the two measures for 1960 (those on the left of the diagram).
- (ii) Some candidates correctly identified the presence of the extreme value 47 per cent and considered its impact or lack of impact on either the range or the median, but it was rare to see fully correct answers for both quantities. Often a vague reference to extreme values (plural) was made, where in fact in 1960 there was only one extreme value. It is important in questions of this sort to give answers that are specific to the context and data provided, rather than to give a generalised 'textbook' type answer.
- (c) This question proved to be challenging, with many candidates incorrectly concluding that 'the diagram could be used' because 'in a stem-and-leaf diagram raw data is not lost', rather than observing that, in listing the data in order on each side of the diagram, the country associated with each percentage has been lost. It is therefore unknown whether there has been an increase in each of the 11 countries.
- (d) Some good clear solutions were seen to this probability question, although some candidates did omit it. Some candidates had solutions of the correct form but did not multiply their product of three probabilities by 3 or did not reduce the denominator as countries were selected.

Question 8

- (a) Most candidates inserted correct price relatives into the table. A few did not recognise that all price relatives in the base year, 2022, should be 100, and some put amounts of dollars rather than price relatives in the table.
- (b) This question was answered well by most candidates. Some gained one out of the two available marks if one of the essential elements, often the years, were missing. Other common errors were 'per cent' missing from 1.5 or 'expenditure' or 'price relative' used instead of 'price'.
- (c) Most candidates made a good attempt at finding the weighted aggregate cost index with many fully correct solutions seen. A small number divided an appropriate sum of products by 100 rather than by the total of the weights, 26.
- (d)(i) Many candidates used the correct method for estimating expenditure using the weighted aggregate cost index, but some did not follow the instruction to give the answer to the specified degree of accuracy. Some candidates divided by rather than multiplied by the weighted aggregate cost index and some candidates omitted this question.
- (ii) The best responses gave reasons, in the context of the hotel, related to a change of quantities, such as 'more electricity was used' or 'fewer repairs were needed'. Common errors included general statements not related to the specific context of the question, such as 'more items were used', or statements about changes to prices that have been accounted for in the price relatives, such as 'the price of electricity increased'.
- (e) This was an unusual question comparing weights in different years. Only the most able candidates noticed that the total of the weights in 2024 had reduced, compared to 2022, and that therefore the relative weight of food had increased. Most candidates saw that the weights associated with food in each year were the same and concluded that the relative weights had not changed.

Question 9

- (a) Many candidates obtained the correct probability of missing the target, although some subtracted each of the given probabilities from 1 and then multiplied the resulting probabilities, giving an answer of 0.6375 rather than 0.6.
- (b)(i) Fully correct probability distributions were seen in the best responses. Some other candidates had a table with the correct winning amounts, but in some cases the probabilities attached to these amounts were just those associated with one throw of the ball. A simple tree diagram might have helped some candidates with finding the probabilities associated with the necessary sequential events. Some candidates found a series of appropriate probabilities but did not gather them into a probability distribution table or did not add together the probabilities of the two ways of winning \$1 or the two ways of winning \$2.
- (ii) Some candidates correctly recognised that they had assumed that the probabilities did not change when throwing the ball a second time. Many candidates, however, simply restated some of the rules of the game as given in the question.
- (iii) Candidates with a fully correct probability distribution in **part (b)(i)** were usually able to find the amount that should be charged for a fair game. Those with some form of probability distribution were usually able to use it appropriately and demonstrate understanding of the concept of a fair game.
- (c)(i) As with **part (b)(i)**, a tree diagram might have helped some candidates to organise the relevant probabilities and understand which should be multiplied and which should be added. Able candidates found some, if not all, of the probabilities required, and the most able multiplied these probabilities by the associated amounts of \$2, \$3 and \$4, to obtain the expected amount that a player will win.
- (ii) Many candidates made a good attempt at the expected profit or loss for the owner, although some quoted a loss when their expectation gave a profit or *vice versa*, and some omitted to state whether the amount found represented a profit or a loss. A small number omitted to multiply their expected amount by 20 players and simply subtracted their expected amount from \$60.

Question 10

- (a) Most candidates used correct statistical language to describe the type of data collected, although a few focused on the purpose of the collection or statistical techniques that were going to be used rather than the type of data. Most candidates recognised the data as quantitative, although a few stated that it was continuous rather than discrete.
- (b) Most candidates obtained all three values that were missing from the table correctly, with just a few getting one or two of them incorrect.
- (c) Fully correct calculations for the seasonal component were seen in many scripts, with the most common error being a positive answer of the correct magnitude rather than a negative answer. A small number found just one of the appropriate differences.
- (d) Correct plots of the centred moving average values were usually seen, with just a small number of candidates misreading the given vertical scale. Many trend lines correctly followed the trend of the plots with a similar number of plots above and below the line. The two most common errors were that some trend lines were not steep enough and that some trend lines had all the plots on or below the trend line. The first case occurred if candidates only followed the trend of the first 4 plots and the second case occurred if candidates simply drew their line through the first and last plots.
- (e) Many candidates correctly added 215 to an appropriate trend line reading. Some candidates incorrectly subtracted, rather than added, 215 and some used their calculated seasonal component for quarter 1, from **part (c)**, together with a reading from their trend line from quarter 2 of 2025.
- (f) This was an unusual question, and many of the more able candidates produced clear explanations for why Jamal's claim could not be made from the data in the table. Explanations ranged from stating the limitations of the seasonal components (only giving information about how far above or below the trend line a quarterly value is, on average) to explaining what additional information would be required to be able to make this claim. Some weaker responses made incorrect comments about numbers of flights not being negative or concluded that Jamal's claim was in fact correct.

STATISTICS

Paper 4040/23 Paper 2
--

Key messages

To be successful in this examination the context of a question should be carefully considered, and interpretation of diagrams, graphs and calculations should always be made with this context in mind. Final answers should also be checked to ensure that they seem sensible in the given context.

Candidates should read questions carefully and provide clearly set out solutions, particularly in questions that begin 'Show ...' and in multistage problems.

General comments

It is important that interpretation of results is made with the context in mind. In **Question 8dii**, for example, it was important to give examples specifically in the context of the hotel and avoid generalised answers that could apply to any context. An example of an appropriate answer was 'the amount of food bought might have changed'. A generalised answer that could apply to any context, such as 'the quantity of items changed', should be avoided.

Checking a final answer in the context of the problem is an important stage in the working. A candidate who had incorrectly divided by 100 rather than by 26, in **Question 8c**, for example, would have reached an answer of the wrong sort of magnitude, that, when considered as representing the overall increase in costs for the hotel, would not have been compatible with the calculated price relatives. Similarly, in **Question 8di**, candidates who divided rather than multiplied by a weighted aggregate cost index that was greater than 100, say, would have reached an estimate for expenditure in 2023 that was lower than the figure for 2022. Considering whether their answer was sensible or not, in the given context, could have alerted them to their error.

Clearly set out solutions were important in many of the questions. **Question 6bii** asked candidates to 'Show ...', for 3 marks. The three required stages in the working, namely, showing how many diesel cars should be in the sample, showing how many diesel cars were in the sample and drawing the appropriate conclusion should be clearly set out.

Some of the multistage problems, such as **Questions 7d** and **10c**, often had clearly set out solutions where candidates who had made errors could receive partial credit because their method was clear. For many candidates, more clearly set out working in the form of probability tree diagrams might have been helpful in **Question 9bi** and **Question 9ci**.

Comments on specific questions

Question 1

Most candidates made a reasonable attempt at this question, although fully correct solutions were rarely seen. Most candidates were able to correctly find the required measures for the most straightforward of the continuous variables, namely the length of time. The other continuous variable, age, had the complication of being in completed years, and some candidates either ignored this aspect or dealt with it incorrectly. Answers for the discrete variable, the number of visitors, were most likely to be incorrect.

Question 2

- (a) (i) Most candidates were able to interpret the sectional bar chart correctly. The small number of errors in this question appeared to come from candidates incorrectly reading values from the vertical scale, but this was very rare.
- (ii) Fully correct solutions to this part were seen in most cases, with just a few candidates finding either the number, rather than the percentage, or finding the percentage who said Internet, rather than Newspapers. Reading the question carefully might have prevented these errors.
- (b) (i) Able candidates were able to give an answer equivalent to 'it appears to be true' because '25 per cent of adults and 31.25 per cent of children use TV and Radio' or because 'the proportion (or percentage) of adults using TV and Radio is smaller than the proportion of children using TV and Radio'. Very few candidates gave a correct comment about the sample size or the reliability of the sample. Many candidates instead suggested that Neo was incorrect based on the numerical values of 10 children and 12 adults, without taking account of the different population sizes.
- (ii) Candidates who had correctly considered percentages in **part (b)(i)** did not always suggest a type of chart that would display percentages, with a dual bar chart being a common incorrect response. Charts that would have allowed Leo to make a comment about likelihood more conveniently would have been, for example, comparative pie charts or a percentage sectional bar chart. Other incorrect responses seen included incomplete answers such as 'percentage bar chart'.

Question 3

- (a) Many candidates demonstrated a good understanding of linear interpolation, with a variety of correct methods leading to a correct median. Those that made an error had generally found the correct class interval for the median journey length.
- (b) It was very common for candidates to assume that the interquartile range would be impacted by the addition of a constant to each piece of data as well as by the multiplication of each piece of data by a constant, with an incorrect answer of \$12.20 being very common. A measure of spread only being impacted by the multiplication by a constant was understood by only a very few candidates.
- (c) Using linear interpolation to find a number of journeys from a given distance proved more difficult than finding the median distance. The most common incorrect answer seen was 95, the number of journeys shorter than, rather than longer than, 18 km. Rereading the question might have prevented this error.

Question 4

- (a) Candidates tended to be more successful in listing the independent events than the mutually exclusive events. The most common error seen in the list of independent events was either one or two of the four pairs omitted, or the inclusion of A and B, the pair of mutually exclusive events, in the wrong list.
- (b) Most candidates were able to score at least some of the marks in this question, with $P(B \text{ and } C)$ being more often found correctly. The most common error was for the two events to be treated as if they were mutually exclusive when it came to finding $P(B \text{ or } C)$, with the subtraction of $P(B \text{ and } C)$ being omitted. A few candidates confused the words 'or' and 'and', incorrectly finding $P(B \cap C)$ for $P(B \text{ or } C)$ and finding $P(B \cup C)$ for $P(B \text{ and } C)$.
- (c) The most common incorrect answer was 4 rather than 5. For the red bag it is known that $P(\text{odd}) = 0.6$ or $3/5$, so there must be at least 5 discs in that bag.

Question 5

- (a) This question was answered well by most candidates. Those who did not score full marks often scored a method mark for having a correct expression for the number of standard deviations below the mean.

- (b) There were many fully correct solutions seen to this part. The most common errors were for a numerical value to be used for one occurrence of the unknown in the equation $\frac{x-23}{6} = \frac{x-30}{9}$, or for different letters to be used, x and y , say, for the two occurrences of the unknown. There were sometimes other expressions/methods, equivalent to the one given above, that were seen.
- (c) This proved to be the most difficult of the three parts to this question, although many fully correct solutions were seen. Some candidates formed a correct initial equation but were unable to solve it correctly. Others tried to find the standard deviation by using the formula $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$, an incorrect method for this question.

Question 6

- (a) In this question it was necessary to adapt the given two-digit random number table to produce equally likely numbers in the range 000 to 359, by dividing the table into 3-digit blocks. Candidates found this question challenging, with most producing a list of two-digit rather than three-digit numbers.
- (b)(i) Many candidates selected a stratified sample in this part, although some selected a sample stratified by both fuel type and gearbox, rather than by gearbox alone. A small number selected a simple random sample, without stratifying.
- (ii) Having selected their sample, many candidates correctly identified which of their sample were diesel cars, usually by listing them. Most also showed the necessary calculation for finding the number that would be in a representative sample, although some candidates did one or other of these without clearly showing both. It is important in a question where candidates are asked to 'show' a particular result, that all the necessary working is clearly laid out. Those that had a correct sample in **part (b)(i)** often provided all the necessary working and drew the correct conclusion from that working.

Question 7

- (a) It was pleasing to see many fully correct descriptions of the key. The most common errors were to only describe the part of the key on the right, so just 17 per cent in 2020, or to incorrectly define the value on the left as 1 per cent in 1960 rather than 10 per cent in 1960.
- (b)(i) Most candidates used the key correctly and found the required medians and ranges. The most common errors came from either slips in locating one or both median values or for errors in finding the two measures for 1960 (those on the left of the diagram).
- (ii) Some candidates correctly identified the presence of the extreme value 47 per cent and considered its impact or lack of impact on either the range or the median, but it was rare to see fully correct answers for both quantities. Often a vague reference to extreme values (plural) was made, where in fact in 1960 there was only one extreme value. It is important in questions of this sort to give answers that are specific to the context and data provided, rather than to give a generalised 'textbook' type answer.
- (c) This question proved to be challenging, with many candidates incorrectly concluding that 'the diagram could be used' because 'in a stem-and-leaf diagram raw data is not lost', rather than observing that, in listing the data in order on each side of the diagram, the country associated with each percentage has been lost. It is therefore unknown whether there has been an increase in each of the 11 countries.
- (d) Some good clear solutions were seen to this probability question, although some candidates did omit it. Some candidates had solutions of the correct form but did not multiply their product of three probabilities by 3 or did not reduce the denominator as countries were selected.

Question 8

- (a) Most candidates inserted correct price relatives into the table. A few did not recognise that all price relatives in the base year, 2022, should be 100, and some put amounts of dollars rather than price relatives in the table.
- (b) This question was answered well by most candidates. Some gained one out of the two available marks if one of the essential elements, often the years, were missing. Other common errors were 'per cent' missing from 1.5 or 'expenditure' or 'price relative' used instead of 'price'.
- (c) Most candidates made a good attempt at finding the weighted aggregate cost index with many fully correct solutions seen. A small number divided an appropriate sum of products by 100 rather than by the total of the weights, 26.
- (d)(i) Many candidates used the correct method for estimating expenditure using the weighted aggregate cost index, but some did not follow the instruction to give the answer to the specified degree of accuracy. Some candidates divided by rather than multiplied by the weighted aggregate cost index and some candidates omitted this question.
- (ii) The best responses gave reasons, in the context of the hotel, related to a change of quantities, such as 'more electricity was used' or 'fewer repairs were needed'. Common errors included general statements not related to the specific context of the question, such as 'more items were used', or statements about changes to prices that have been accounted for in the price relatives, such as 'the price of electricity increased'.
- (e) This was an unusual question comparing weights in different years. Only the most able candidates noticed that the total of the weights in 2024 had reduced, compared to 2022, and that therefore the relative weight of food had increased. Most candidates saw that the weights associated with food in each year were the same and concluded that the relative weights had not changed.

Question 9

- (a) Many candidates obtained the correct probability of missing the target, although some subtracted each of the given probabilities from 1 and then multiplied the resulting probabilities, giving an answer of 0.6375 rather than 0.6.
- (b)(i) Fully correct probability distributions were seen in the best responses. Some other candidates had a table with the correct winning amounts, but in some cases the probabilities attached to these amounts were just those associated with one throw of the ball. A simple tree diagram might have helped some candidates with finding the probabilities associated with the necessary sequential events. Some candidates found a series of appropriate probabilities but did not gather them into a probability distribution table or did not add together the probabilities of the two ways of winning \$1 or the two ways of winning \$2.
- (ii) Some candidates correctly recognised that they had assumed that the probabilities did not change when throwing the ball a second time. Many candidates, however, simply restated some of the rules of the game as given in the question.
- (iii) Candidates with a fully correct probability distribution in **part (b)(i)** were usually able to find the amount that should be charged for a fair game. Those with some form of probability distribution were usually able to use it appropriately and demonstrate understanding of the concept of a fair game.
- (c)(i) As with **part (b)(i)**, a tree diagram might have helped some candidates to organise the relevant probabilities and understand which should be multiplied and which should be added. Able candidates found some, if not all, of the probabilities required, and the most able multiplied these probabilities by the associated amounts of \$2, \$3 and \$4, to obtain the expected amount that a player will win.
- (ii) Many candidates made a good attempt at the expected profit or loss for the owner, although some quoted a loss when their expectation gave a profit or *vice versa*, and some omitted to state whether the amount found represented a profit or a loss. A small number omitted to multiply their expected amount by 20 players and simply subtracted their expected amount from \$60.

Question 10

- (a) Most candidates used correct statistical language to describe the type of data collected, although a few focused on the purpose of the collection or statistical techniques that were going to be used rather than the type of data. Most candidates recognised the data as quantitative, although a few stated that it was continuous rather than discrete.
- (b) Most candidates obtained all three values that were missing from the table correctly, with just a few getting one or two of them incorrect.
- (c) Fully correct calculations for the seasonal component were seen in many scripts, with the most common error being a positive answer of the correct magnitude rather than a negative answer. A small number found just one of the appropriate differences.
- (d) Correct plots of the centred moving average values were usually seen, with just a small number of candidates misreading the given vertical scale. Many trend lines correctly followed the trend of the plots with a similar number of plots above and below the line. The two most common errors were that some trend lines were not steep enough and that some trend lines had all the plots on or below the trend line. The first case occurred if candidates only followed the trend of the first 4 plots and the second case occurred if candidates simply drew their line through the first and last plots.
- (e) Many candidates correctly added 215 to an appropriate trend line reading. Some candidates incorrectly subtracted, rather than added, 215 and some used their calculated seasonal component for quarter 1, from **part (c)**, together with a reading from their trend line from quarter 2 of 2025.
- (f) This was an unusual question, and many of the more able candidates produced clear explanations for why Jamal's claim could not be made from the data in the table. Explanations ranged from stating the limitations of the seasonal components (only giving information about how far above or below the trend line a quarterly value is, on average) to explaining what additional information would be required to be able to make this claim. Some weaker responses made incorrect comments about numbers of flights not being negative or concluded that Jamal's claim was in fact correct.