STATISTICS

Paper 4040/12
Paper 1

Key messages

It is important always to read carefully and appreciate fully the practical situation described in a question and the nature of the given data.

After obtaining a calculated result it is worth pausing to consider whether or not the result seems possible and reasonable for the practical situation and data of the question.

It is just as important to develop the ability to interpret the results of calculations as it is to carry them out accurately.

General comments

The overall standard of work involving calculations of a routine nature was good. This was particularly true of calculations on mean and standard deviation, line of best fit and standardised rates. There were good answers on some of the writing parts of questions (see **Question 8** below), but also more limited answers on interpreting the differences shown by diagrams and calculated measures (see **Question 10** below). Performance on the probability questions was generally not good (see **Questions 6** and **9** below). On the topics of Venn diagrams and histograms, performance varied a lot.

It has been emphasised repeatedly in these reports that Statistics is a subject which is applied to real-life situations and that candidates should try to be aware of whether or not their answers are reasonable for the situation of the question. There was a question on this paper where some practically impossible answers were presented, answers which should have given the candidate pause for thought about what they must have done incorrectly (see **Question 10** below).

There were a few instances where candidates did their graphs on locally supplied graph paper instead of, as required, on the question paper.

Comments on specific questions

Question 1

The statistical terms were generally well understood. Errors were most common in **part (e)**.

Question 2

Almost all candidates were able to interpret the value 2 in the diagram. Answers to the other parts of the question were more mixed. Some candidates still demonstrate limited understanding of what the different regions of a Venn diagram represent.

Question 3

Many fully correct answers to **part (a)** were seen. Calculations were often carried out with maximum accuracy, in particular using the exact value of the mean when finding the standard deviation instead of approximating it, as has sometimes been the case in the past. There were far fewer good answers to **part (b)**. Limited understanding was shown of why it is desirable that a standard deviation in such a situation should be as small as possible, and why therefore there should be no lower limit placed upon it.

Question 4

It seems to be commonly understood now that in a histogram not all class frequencies can be found simply by reading column heights. But the principle of frequency being proportional to column area is still not always applied properly from the labelling of the scale on the vertical axis. Thus, for example, multiples of the correct answers were sometimes seen in **part (a)**. Although the working involved in such questions is minimal, candidates are still advised to show it; it is impossible for an Examiner to award method marks if an answer is simply stated without working and it is incorrect. It seems as though in answering **part (b)** many candidates had not read properly the information preceding it, and that it was not '3 hours up to 6 hours' that had to be used when reading the histogram.

Question 5

Performance on this question was generally not good. In particular there was quite often confusion between measures of central tendency and measures of dispersion, an absolutely fundamental distinction in this subject. It was not unusual in **part**s **(a)** and **(b)** to see measures of central tendency presented. When this happened the candidate was usually destined to earn no marks on the question. Candidates who performed well knew the difference between these two classes of measures, and understood clearly the problem created by the open class in determining whether or not they could be calculated exactly in this case. Amongst candidates who performed well it was usually in **part (d)** where limitations were seen, there being limited appreciation that it was both the 21st and the 22nd items which had to identified in the ordered distribution to find the new median.

Question 6

Most candidates appreciated the fact that the probabilities here were 'without replacement' probabilities, and there were many correct answers to **part (a)**. A common limitation in **part (b)** was that very often only three cases were counted instead of the correct six, though such answers usually earned some credit. It was in **part (c)** that many candidates created severe difficulties for themselves. This was almost universally the result of the players not being re-categorised as strong/not strong, the relevant characteristics for the question, which leads to a limited number of possible cases of interest. In retaining the original three categories there were so many cases to consider that much needless working was seen and almost inevitably something was missed. The few candidates who re-categorised appropriately usually earned at least some credit with very little working.

Question 7

The first three parts were answered well. Although the lower semi-average was often found correctly in **part (b)**, candidates should be advised to use the two given averages in such a question when finding the line of best fit, and not the one they have been asked to calculate, in case the latter has been calculated incorrectly. The difference between young and mature trees was well understood in **part (d)**, and some well-drawn lines 'by eye' were seen in **part (e)**. In **part (f)**, whilst almost all candidates knew how to use the line they had drawn, very often the labelling of the vertical scale of the grid was ignored: often an answer of, say, 3.7 might be presented in the answer space, instead of 3700.

It was in **part (g)** where answers were most limited. Much space was needlessly occupied with percentage work, often incorrect, where brief reflection might have resulted in the appreciation that if four fifths of the mass has been lost, then the 1000 kg represents one fifth of the original mass, 5000 kg. Candidates who realised this were able to obtain the correct answer very quickly from the line drawn in **part (e)**. Any use of the equation from **part (c)** was invalid, as this equation did not relate to mature trees only.

Question 8

Good general understanding was shown in explaining the negative correlation in **part (a)**. Whilst the majority of candidates also probably understood in **part (b)** why the standardised injury rate for Construction was the highest, many answers were incomplete because they did not refer to all job tenure groups. Good computational skills were shown in **part (c)(i)** where many fully correct answers were seen, followed by a correctly stated conclusion in **part (c)(ii)**.

Answers to **part (d)** varied more in quality. Candidates who realised that the relevant job tenure group injury rates from the first table had to be used with each group population in the second table were usually successful. Those using column totals from the first table or data in the second table alone produced work

which was totally invalid. In **part (e)** some allowance was made for candidates using an incorrect value from **part (d)**.

Question 9

Answers to **part (b)** varied in quality. Many well-drawn correct graphs were seen, but there were many also with a major error: the plotting of the points at either the mid-points or lower boundaries of the class intervals, rather than at the upper boundaries. Candidates need to be aware of the importance of correct plotting, as incorrect plotting severely affects later answers when reading values from the graph.

Parts (c)(i) and **(c)(ii)** were well done, but the remaining parts less so. In **part (c)(iii)** it was essential for the candidate to recognise that the 44 km/h was not itself to be used in a calculation, but the cumulative frequency corresponding to this speed. In **part (d)** only the candidates who saw that the question involved the 26 hunting runs with speeds of 50 km/h or more were able to gain credit. Answers to the probability question in **part (e)** tended to be limited in two ways: there was not always recognition that there were two cases involved; and it was quite common for 'without replacement' probabilities to be used.

Question 10

Almost all candidates were able to draw accurately a pie chart with the correct angles. However, to obtain full credit in **part (a)** it was necessary to use the different totals of visitors to the two attractions to draw a chart with a radius which reflected this difference. Far too many candidates ignored the fact that there had been more visitors to the Museum than to the Aquarium and drew the second chart with a radius much the same as the one given. To obtain credit in **part (b)**, when comparing the charts, the observations had to be expressed in terms of proportions (or fractions or percentages). Many candidates answered in terms of absolute numbers, making statements which not only conflicted with the data in the given table, but making comparisons that could in any case easily be seen in the original data.

Many correct answers were seen for the mean age calculations in **part (c)**, but in some cases answers were presented which were absolutely impossible for these distributions. In all statistical work, candidates should try to be aware of what constitutes a reasonable answer in a given practical situation. Here, for example, for the given data, it should be obvious that the mean ages could not possibly be less than 20 years or more than 80 years. Although having correct answers for the mean ages, few candidates were able to draw an appropriate conclusion from their results, because apparently the information given before the second table had not been read properly.

There were many fully correct answers to **parts (d)** and **(e)**.

STATISTICS

Paper 4040/13
Paper 1

Key messages

It is important always to read carefully and appreciate fully the practical situation described in a question and the nature of the given data.

After obtaining a calculated result it is worth pausing to consider whether or not the result seems possible and reasonable for the practical situation and data of the question.

It is just as important to develop the ability to interpret the results of calculations as it is to carry them out accurately.

General comments

The overall standard of work involving calculations of a routine nature was good. This was particularly true of calculations on mean and standard deviation, line of best fit and standardised rates. There were good answers on some of the writing parts of questions (see **Question 8** below), but also more limited answers on interpreting the differences shown by diagrams and calculated measures (see **Question 10** below). Performance on the probability questions was generally not good (see **Questions 6** and **9** below). On the topics of Venn diagrams and histograms, performance varied a lot.

It has been emphasised repeatedly in these reports that Statistics is a subject which is applied to real-life situations and that candidates should try to be aware of whether or not their answers are reasonable for the situation of the question. There was a question on this paper where some practically impossible answers were presented, answers which should have given the candidate pause for thought about what they must have done incorrectly (see **Question 10** below).

There were a few instances where candidates did their graphs on locally supplied graph paper instead of, as required, on the question paper.

Comments on specific questions

Question 1

The statistical terms were generally well understood. Errors were most common in **part (e)**.

Question 2

Almost all candidates were able to interpret the value 2 in the diagram. Answers to the other parts of the question were more mixed. Some candidates still demonstrate limited understanding of what the different regions of a Venn diagram represent.

Question 3

Many fully correct answers to **part (a)** were seen. Calculations were often carried out with maximum accuracy, in particular using the exact value of the mean when finding the standard deviation instead of approximating it, as has sometimes been the case in the past. There were far fewer good answers to **part (b)**. Limited understanding was shown of why it is desirable that a standard deviation in such a situation should be as small as possible, and why therefore there should be no lower limit placed upon it.

Question 4

It seems to be commonly understood now that in a histogram not all class frequencies can be found simply by reading column heights. But the principle of frequency being proportional to column area is still not always applied properly from the labelling of the scale on the vertical axis. Thus, for example, multiples of the correct answers were sometimes seen in **part (a)**. Although the working involved in such questions is minimal, candidates are still advised to show it; it is impossible for an Examiner to award method marks if an answer is simply stated without working and it is incorrect. It seems as though in answering **part (b)** many candidates had not read properly the information preceding it, and that it was not '3 hours up to 6 hours' that had to be used when reading the histogram.

Question 5

Performance on this question was generally not good. In particular there was quite often confusion between measures of central tendency and measures of dispersion, an absolutely fundamental distinction in this subject. It was not unusual in **part**s **(a)** and **(b)** to see measures of central tendency presented. When this happened the candidate was usually destined to earn no marks on the question. Candidates who performed well knew the difference between these two classes of measures, and understood clearly the problem created by the open class in determining whether or not they could be calculated exactly in this case. Amongst candidates who performed well it was usually in **part (d)** where limitations were seen, there being limited appreciation that it was both the 21st and the 22nd items which had to identified in the ordered distribution to find the new median.

Question 6

Most candidates appreciated the fact that the probabilities here were 'without replacement' probabilities, and there were many correct answers to **part (a)**. A common limitation in **part (b)** was that very often only three cases were counted instead of the correct six, though such answers usually earned some credit. It was in **part (c)** that many candidates created severe difficulties for themselves. This was almost universally the result of the players not being re-categorised as strong/not strong, the relevant characteristics for the question, which leads to a limited number of possible cases of interest. In retaining the original three categories there were so many cases to consider that much needless working was seen and almost inevitably something was missed. The few candidates who re-categorised appropriately usually earned at least some credit with very little working.

Question 7

The first three parts were answered well. Although the lower semi-average was often found correctly in **part (b)**, candidates should be advised to use the two given averages in such a question when finding the line of best fit, and not the one they have been asked to calculate, in case the latter has been calculated incorrectly. The difference between young and mature trees was well understood in **part (d)**, and some well-drawn lines 'by eye' were seen in **part (e)**. In **part (f)**, whilst almost all candidates knew how to use the line they had drawn, very often the labelling of the vertical scale of the grid was ignored: often an answer of, say, 3.7 might be presented in the answer space, instead of 3700.

It was in **part (g)** where answers were most limited. Much space was needlessly occupied with percentage work, often incorrect, where brief reflection might have resulted in the appreciation that if four fifths of the mass has been lost, then the 1000 kg represents one fifth of the original mass, 5000 kg. Candidates who realised this were able to obtain the correct answer very quickly from the line drawn in **part (e)**. Any use of the equation from **part (c)** was invalid, as this equation did not relate to mature trees only.

Question 8

Good general understanding was shown in explaining the negative correlation in **part (a)**. Whilst the majority of candidates also probably understood in **part (b)** why the standardised injury rate for Construction was the highest, many answers were incomplete because they did not refer to all job tenure groups. Good computational skills were shown in **part (c)(i)** where many fully correct answers were seen, followed by a correctly stated conclusion in **part (c)(ii)**.

Answers to **part (d)** varied more in quality. Candidates who realised that the relevant job tenure group injury rates from the first table had to be used with each group population in the second table were usually successful. Those using column totals from the first table or data in the second table alone produced work

which was totally invalid. In **part (e)** some allowance was made for candidates using an incorrect value from **part (d)**.

Question 9

Answers to **part (b)** varied in quality. Many well-drawn correct graphs were seen, but there were many also with a major error: the plotting of the points at either the mid-points or lower boundaries of the class intervals, rather than at the upper boundaries. Candidates need to be aware of the importance of correct plotting, as incorrect plotting severely affects later answers when reading values from the graph.

Parts (c)(i) and **(c)(ii)** were well done, but the remaining parts less so. In **part (c)(iii)** it was essential for the candidate to recognise that the 44 km/h was not itself to be used in a calculation, but the cumulative frequency corresponding to this speed. In **part (d)** only the candidates who saw that the question involved the 26 hunting runs with speeds of 50 km/h or more were able to gain credit. Answers to the probability question in **part (e)** tended to be limited in two ways: there was not always recognition that there were two cases involved; and it was quite common for 'without replacement' probabilities to be used.

Question 10

Almost all candidates were able to draw accurately a pie chart with the correct angles. However, to obtain full credit in **part (a)** it was necessary to use the different totals of visitors to the two attractions to draw a chart with a radius which reflected this difference. Far too many candidates ignored the fact that there had been more visitors to the Museum than to the Aquarium and drew the second chart with a radius much the same as the one given. To obtain credit in **part (b)**, when comparing the charts, the observations had to be expressed in terms of proportions (or fractions or percentages). Many candidates answered in terms of absolute numbers, making statements which not only conflicted with the data in the given table, but making comparisons that could in any case easily be seen in the original data.

Many correct answers were seen for the mean age calculations in **part (c)**, but in some cases answers were presented which were absolutely impossible for these distributions. In all statistical work, candidates should try to be aware of what constitutes a reasonable answer in a given practical situation. Here, for example, for the given data, it should be obvious that the mean ages could not possibly be less than 20 years or more than 80 years. Although having correct answers for the mean ages, few candidates were able to draw an appropriate conclusion from their results, because apparently the information given before the second table had not been read properly.

There were many fully correct answers to **parts (d)** and **(e)**.

STATISTICS

Paper 4040/22
Paper 2

Key message

In this examination candidates need to demonstrate the ability to make decisions about the most appropriate statistical techniques to employ. They also need to be able to employ those techniques accurately. For pictorial representations this means plotting accurately and labelling appropriately, and for statistical calculations this means setting out the work clearly to show all the stages of the calculation. The results of this statistical analysis must then be interpreted in the context of the situations presented. Thus the full cycle of statistical enquiry is explored within this examination, from planning and data collection, to presentation and analysis, and finally to the interpretation of that analysis.

General comments

In terms of planning some candidates had difficulty choosing an appropriate multiple bar chart in **Question 1(a)**. This lead to difficulties in answering **Question 1(b)** and highlights the need to consider carefully the purpose of a statistical representation when deciding which one to use.

Presentation of data was examined with a sectional bar chart, a stem-and-leaf diagram, a time series graph and a box-and-whisker diagram. Where a scale was provided most candidates were able to use it accurately, but labelling was sometimes either missing or incomplete. For example, the key for the back-to-back stem-and-leaf diagram in **Question 4** was sometimes missing or only explained the figures on the right of the diagram.

Questions that involved statistical calculations were usually well presented, and it was particularly pleasing to see some well-structured solutions to **Question 7(b)**. It is very important that working is clearly set out so that partial credit for incomplete or incorrect solutions can be awarded. It is also important that working is clearly presented when an answer is given in the question, as was the case in **Question 5(a)**.

As has already been mentioned, some candidates had difficulty with the interpretation required in **Question 1(b)**, particularly if they had made an inappropriate choice of diagram in **part (a)**. Elsewhere in the paper where interpretation was required, such as in **Question 4(b)**, where analysis of the pictorial representation as well as the mean was needed, many candidates had difficulty. In **Question 8(e)** it was important to provide interpretation in the context of the question. It was not therefore sufficient simply to say, for example, that the median for cars was smaller than that of pickup trucks. Rather, it was necessary to interpret this fact and explain that, therefore, the cars used less fuel on average.

Candidates would benefit from rereading questions once they have completed them to check that they have answered them fully. For example, in **Question 8(b)**, some candidates gave the number of cars with a fuel consumption greater than 9 *l*/100 km, rather than the percentage; and in **Question 9(d)(i)**, some candidates did not give their answer to the requested degree of accuracy.

Comments on specific questions

Question 1

In this question data relating to the population of a town was presented in a multiple bar chart. Candidates were asked to present the same data in a different chart in order check Azeeb's claim. It was therefore important, in **part (a)**, to consider this purpose when deciding how to organise the sectional bar chart. Azeeb's claim was about changes in population over time, and therefore it was necessary to present bars for each year to allow for appropriate comparisons. Azeeb's claim was also about the overall population of the

town and, therefore, each of the three bars needed to comprise the populations from each age group. Common errors were for the three bars each to represent an age group or for a percentage sectional bar chart to be drawn for each year. Neither of these representations allowed the candidates to check Azeeb's claim.

Those candidates that drew an appropriate chart usually provided the necessary labelling, although 'thousands' was sometimes missing from the vertical axis. The scales used were usually appropriate with accurately drawn bar heights.

For the two marks in **part (b)**, candidates needed to consider the two aspects of Azeeb's claim, namely whether or not there had been an increase in the overall population and whether or not that increase was steady. Candidates who had not drawn an appropriate chart were often unable to comment on the second part of his claim, but could use the original chart to see that there had been in increase in the overall population. Those that had drawn an appropriate diagram were in a position also to compare the overall population increase from 1991 to 2001 with that from 2001 and 2011. It was equally acceptable to consider this as 'not steady' because the increase from 2001 to 2011 was greater than that from 1991 to 2001, or to consider it as 'steady' because there was an overall increase in each period. It was not, however, sufficient simply to state 'steady' or 'not steady' without giving a reason for that decision.

Question 2

In **part (a)** some candidates provided a general definition of a population, rather than describing the population in this situation. Some candidates seemed to confuse the population with the sample.

Most candidates correctly found a simple random sample of the candidates, in **part (b)**, with just a small number repeating the 21.

Errors in **part (c)** were more common, with some candidates appearing to be confused about the nature of a systematic sample, picking what appeared to be another simple random sample from the new table. Some candidates correctly calculated the interval size of 5, but then chose every fifth value from the random number table. Other candidates scored some of the marks using a systematic sample but with the wrong interval size, often 6, or with the correct interval size, but the wrong starting value.

Question 3

It was quite rare to see four correct answers in **part (a)**. When comparing variable *W* with variable *V*, some candidates appeared to think that each value had increased by 10 rather than 9, and thus a new mean of 12.96, rather than 11.96, was quite commonly seen. The most commonly seen correct answer was for the unchanged standard deviation of *W*. For variable *X*, it was quite common to see the standard deviation incorrectly left unchanged.

Candidates were often more successful with **part (b)**, usually getting at least two of the four values correct. The most common error was for the class interval to be given as 4 rather than 5.

Part (c) of this question was challenging, with most candidates not using the mid-points to estimate the mean and standard deviation for *Y*. It was common to see the standard deviation left unchanged.

Question 4

Most candidates were able to produce a stem-and-leaf diagram in **part (a)** with appropriately ordered leaves. Marks were sometimes lost for leaves that were not correctly aligned. For example, for Takala, the 3 leaves representing messages with 12, 13 and 17 characters sometimes extended as far as the 4 leaves representing messages with 1, 2, 9 and 9 characters. Other errors seen included the omission of '5' from the stem or an incomplete key which only explained the values to the right of the stem.

In **part (b)**, many candidates incorrectly concluded that Takala was correct, without considering the effect that the extreme values of 40 and 65 had had on her mean. Looking at the stem-and-leaf diagram it can be seen that, if the extreme values are ignored, Takala has the shorter messages.

Question 5

In **part (a)**, candidates who realised that they needed first to find the number of students usually went on to find the standard deviation successfully, showing sufficient working. Those that did not often incorrectly used 2046 for the number of students in the formula for the standard deviation.

Candidates who set up a correct equation with a repeated variable, in **part (b)**, usually solved it correctly. Some, however, used two different variables in an otherwise correct looking equation or tried to use 2046 to replace one occurrence of the variable.

Question 6

Some fully correct solutions were seen, but many candidates ignored the information at the start of the question that the team played 80 per cent of its matches in its red kit. Those that did use this information sometimes treated it as 80 matches rather than 80 per cent of the matches. Those candidates that were successful usually provided well organised solutions.

Question 7

Most candidates produced an accurate time series graph in **part (a)**, with the plots correctly joined with straight line segments.

It was very pleasing to see some very well organised solutions to **part (b)**. This question was less structured than many in the recent past on this topic, but candidates coped well with this lack of structure, often producing a table of their own, including a column for totals, before inserting values into the table provided.

Many candidates correctly used their results to find the required seasonal component in **part (c)**. Moving average values were usually plotted accurately in **part (d)**, and trend lines were often appropriate. A few candidates drew a trend line that was clearly too steep, with the three moving averages on the left explicitly below their trend line and the three moving averages on the right explicitly above their trend line.

As has been seen in the past, some candidates tried to just use the trend line to make the estimate in **part (e)**, without using the seasonal component. Those that used the seasonal component usually did so correctly.

Many candidates struggled with **part (f)**. Candidates could state either that they were assuming the trend continued or that the seasonal variation continued. Some candidates resorted to leaving this question blank and some others described their method for reaching the answer, rather than stating the assumption that they were making.

Question 8

Part (a) was well answered by most candidates, with the most common error being the length described as discrete rather than continuous.

Parts (b) and **(c)** both required the use of linear interpolation, but far more candidates were successful with **part (c)** than **part (b)**. There were a number of different, correct approaches used in both parts, but some candidates seemed to be trying to employ the sorts of techniques that would work in **part (c)** when they approached **part (b)**. Those that applied a correct technique in **part (b)** sometimes forgot to give their answer as a percentage or found the percentage of cars with a fuel consumption less, rather than more, than 9 *l*/100 km. Many fully correct solutions were seen to **part (c)**.

Most candidates used the information provided together with their answer to the previous part to draw an accurate box-and-whisker diagram in **part (d)**. A small number of candidates plotted the median exactly halfway between the lower and upper quartile, rather than at the position of their median from **part (c)**.

Most candidates were able to give at least one correct comparison of the fuel consumptions of the cars and the pickup trucks in **part (e)**. To score both marks consideration should be given to comparing both the central tendency and a measure of spread, giving that comparison in the context of the question. Some candidates, however, effectively gave the same comparison twice: for example, stating, firstly, that the fuel consumption of the cars is generally lower and then, secondly, that the fuel consumption of the pickup trucks is generally higher; or, firstly, comparing the two ranges and concluding that the variation in fuel consumption

is equal for the two types of vehicle and, secondly, comparing the two interquartile ranges and drawing the same conclusion.

Question 9

In **part (a),** quite a large number of candidates interpreted the price relative as telling them that there was no change in the 'expenditure' rather than that was no change in the 'cost'. This error has also been seen in previous years, where candidates assume that a price relative is giving information about expenditure rather than realising that it is independent of the amount used and is therefore only giving information about the cost. For full marks it was also necessary to explain that the figure related to electricity between 2016 and 2019.

In **part (b),** candidates were shown a calculation that did not make any use of weights, and were asked why it was not likely to produce a good estimate. It was surprisingly rare to see answers that referred to the lack of weights, with some candidates simply saying, 'the weights may have changed', an answer which appeared to assume that weights had in fact been used. It was extremely rare to see candidates also noting that the expenditures for each category (given at the start of the question), which could have been used for weights, were very different from each other and therefore their omission would have a large effect on the result.

Many candidates correctly calculated the price relative for electricity in 2020 using the unit costs provided in **part (c)**. It was quite common, however, to see the price relative for other costs given as 101 rather than 99.

Most candidates used a correct method to find the weighted aggregate cost index in **part (d)(i)**, although it was quite common to see this given to three significant figures rather than to the one decimal place requested in the question. A surprisingly common error seen in **part (d)(ii)** was for candidates to recalculate a mean price relative for 2020 using the method that they had been asked to criticise in **part (b)**. However, many fully correct solutions were also seen.

In **part (e)**, many candidates correctly identified *B* and *D* as the two explanations that could not be used to explain the inaccurate result, but only about half of these were able to provide the reason, namely that these figures had already been accounted for in the calculation.

Question 10

Many candidates in **part (a)** demonstrated some understanding of mutually exclusive and independent events, but it was rare to see a fully correct answer. The most commonly seen partially correct answer was *C*&*D* as a correct pair of mutually exclusive events and *A*&*B* as the correct pair of independent events. Missing often, therefore, was *A*&*D* and *B*&*C* from the list of mutually exclusive events.

Many candidates did not seem to be using their answers to **part (a)** when it came to answering **part (b)**. So, for example, it was common to see non-zero answers to **part (b)(i)**, even from candidates who had correctly identified *A*&*D* as a pair of mutually exclusive in **part (a)**. More candidates were successful with **part (b)(ii)** than the rest of **part (b)**, although some candidates did add rather than multiply the required probabilities. In **part (b)(iii)** many candidates correctly found P(*C*), but instead of simply adding this to P(*B*) they often also subtracted the product of the probabilities, even if they had identified *B*&*C* as a pair of mutually exclusive in **part (a)**.

Part (c) of this question was often well done, with those that did not score full marks often scoring partial credit for recognising, and demonstrating in their calculation, that selecting one white counter can happen in three ways and/or for recognising that the counters are not replaced, demonstrated by having $n(n-1)(n-2)$ in the denominator of their product.

Part (d) was a challenging question. Some candidates correctly identified that to achieve the given aim of each container having counters of only one colour, three black counters must be removed from the bag and then two white counters must be removed from the box. Some candidates were able to find the probability of removing three black counters from the bag correctly. A common error was to continue to assume that the total number of counters in the box had remained at 7 rather than that it had increased to 10. A further common error was for the probability of selecting three black counters from the bag to be added to, rather than multiplied by, the probability of selecting two white counters from the box. Some candidates did score full marks for this question, but also it was left blank in quite a large number of cases.

STATISTICS

Paper 4040/23
Paper 2

Key message

In this examination candidates need to demonstrate the ability to make decisions about the most appropriate statistical techniques to employ. They also need to be able to employ those techniques accurately. For pictorial representations this means plotting accurately and labelling appropriately, and for statistical calculations this means setting out the work clearly to show all the stages of the calculation. The results of this statistical analysis must then be interpreted in the context of the situations presented. Thus the full cycle of statistical enquiry is explored within this examination, from planning and data collection, to presentation and analysis, and finally to the interpretation of that analysis.

General comments

In terms of planning some candidates had difficulty choosing an appropriate multiple bar chart in **Question 1(a)**. This lead to difficulties in answering **Question 1(b)** and highlights the need to consider carefully the purpose of a statistical representation when deciding which one to use.

Presentation of data was examined with a sectional bar chart, a stem-and-leaf diagram, a time series graph and a box-and-whisker diagram. Where a scale was provided most candidates were able to use it accurately, but labelling was sometimes either missing or incomplete. For example, the key for the back-to-back stem-and-leaf diagram in **Question 4** was sometimes missing or only explained the figures on the right of the diagram.

Questions that involved statistical calculations were usually well presented, and it was particularly pleasing to see some well-structured solutions to **Question 7(b)**. It is very important that working is clearly set out so that partial credit for incomplete or incorrect solutions can be awarded. It is also important that working is clearly presented when an answer is given in the question, as was the case in **Question 5(a)**.

As has already been mentioned, some candidates had difficulty with the interpretation required in **Question 1(b)**, particularly if they had made an inappropriate choice of diagram in **part (a)**. Elsewhere in the paper where interpretation was required, such as in **Question 4(b)**, where analysis of the pictorial representation as well as the mean was needed, many candidates had difficulty. In **Question 8(e)** it was important to provide interpretation in the context of the question. It was not therefore sufficient simply to say, for example, that the median for cars was smaller than that of pickup trucks. Rather, it was necessary to interpret this fact and explain that, therefore, the cars used less fuel on average.

Candidates would benefit from rereading questions once they have completed them to check that they have answered them fully. For example, in **Question 8(b)**, some candidates gave the number of cars with a fuel consumption greater than 9 *l*/100 km, rather than the percentage; and in **Question 9(d)(i)**, some candidates did not give their answer to the requested degree of accuracy.

Comments on specific questions

Question 1

In this question data relating to the population of a town was presented in a multiple bar chart. Candidates were asked to present the same data in a different chart in order check Azeeb's claim. It was therefore important, in **part (a)**, to consider this purpose when deciding how to organise the sectional bar chart. Azeeb's claim was about changes in population over time, and therefore it was necessary to present bars for each year to allow for appropriate comparisons. Azeeb's claim was also about the overall population of the

town and, therefore, each of the three bars needed to comprise the populations from each age group. Common errors were for the three bars each to represent an age group or for a percentage sectional bar chart to be drawn for each year. Neither of these representations allowed the candidates to check Azeeb's claim.

Those candidates that drew an appropriate chart usually provided the necessary labelling, although 'thousands' was sometimes missing from the vertical axis. The scales used were usually appropriate with accurately drawn bar heights.

For the two marks in **part (b)**, candidates needed to consider the two aspects of Azeeb's claim, namely whether or not there had been an increase in the overall population and whether or not that increase was steady. Candidates who had not drawn an appropriate chart were often unable to comment on the second part of his claim, but could use the original chart to see that there had been in increase in the overall population. Those that had drawn an appropriate diagram were in a position also to compare the overall population increase from 1991 to 2001 with that from 2001 and 2011. It was equally acceptable to consider this as 'not steady' because the increase from 2001 to 2011 was greater than that from 1991 to 2001, or to consider it as 'steady' because there was an overall increase in each period. It was not, however, sufficient simply to state 'steady' or 'not steady' without giving a reason for that decision.

Question 2

In **part (a)** some candidates provided a general definition of a population, rather than describing the population in this situation. Some candidates seemed to confuse the population with the sample.

Most candidates correctly found a simple random sample of the candidates, in **part (b)**, with just a small number repeating the 21.

Errors in **part (c)** were more common, with some candidates appearing to be confused about the nature of a systematic sample, picking what appeared to be another simple random sample from the new table. Some candidates correctly calculated the interval size of 5, but then chose every fifth value from the random number table. Other candidates scored some of the marks using a systematic sample but with the wrong interval size, often 6, or with the correct interval size, but the wrong starting value.

Question 3

It was quite rare to see four correct answers in **part (a)**. When comparing variable *W* with variable *V*, some candidates appeared to think that each value had increased by 10 rather than 9, and thus a new mean of 12.96, rather than 11.96, was quite commonly seen. The most commonly seen correct answer was for the unchanged standard deviation of *W*. For variable *X*, it was quite common to see the standard deviation incorrectly left unchanged.

Candidates were often more successful with **part (b)**, usually getting at least two of the four values correct. The most common error was for the class interval to be given as 4 rather than 5.

Part (c) of this question was challenging, with most candidates not using the mid-points to estimate the mean and standard deviation for *Y*. It was common to see the standard deviation left unchanged.

Question 4

Most candidates were able to produce a stem-and-leaf diagram in **part (a)** with appropriately ordered leaves. Marks were sometimes lost for leaves that were not correctly aligned. For example, for Takala, the 3 leaves representing messages with 12, 13 and 17 characters sometimes extended as far as the 4 leaves representing messages with 1, 2, 9 and 9 characters. Other errors seen included the omission of '5' from the stem or an incomplete key which only explained the values to the right of the stem.

In **part (b)**, many candidates incorrectly concluded that Takala was correct, without considering the effect that the extreme values of 40 and 65 had had on her mean. Looking at the stem-and-leaf diagram it can be seen that, if the extreme values are ignored, Takala has the shorter messages.

Question 5

In **part (a)**, candidates who realised that they needed first to find the number of students usually went on to find the standard deviation successfully, showing sufficient working. Those that did not often incorrectly used 2046 for the number of students in the formula for the standard deviation.

Candidates who set up a correct equation with a repeated variable, in **part (b)**, usually solved it correctly. Some, however, used two different variables in an otherwise correct looking equation or tried to use 2046 to replace one occurrence of the variable.

Question 6

Some fully correct solutions were seen, but many candidates ignored the information at the start of the question that the team played 80 per cent of its matches in its red kit. Those that did use this information sometimes treated it as 80 matches rather than 80 per cent of the matches. Those candidates that were successful usually provided well organised solutions.

Question 7

Most candidates produced an accurate time series graph in **part (a)**, with the plots correctly joined with straight line segments.

It was very pleasing to see some very well organised solutions to **part (b)**. This question was less structured than many in the recent past on this topic, but candidates coped well with this lack of structure, often producing a table of their own, including a column for totals, before inserting values into the table provided.

Many candidates correctly used their results to find the required seasonal component in **part (c)**. Moving average values were usually plotted accurately in **part (d)**, and trend lines were often appropriate. A few candidates drew a trend line that was clearly too steep, with the three moving averages on the left explicitly below their trend line and the three moving averages on the right explicitly above their trend line.

As has been seen in the past, some candidates tried to just use the trend line to make the estimate in **part (e)**, without using the seasonal component. Those that used the seasonal component usually did so correctly.

Many candidates struggled with **part (f)**. Candidates could state either that they were assuming the trend continued or that the seasonal variation continued. Some candidates resorted to leaving this question blank and some others described their method for reaching the answer, rather than stating the assumption that they were making.

Question 8

Part (a) was well answered by most candidates, with the most common error being the length described as discrete rather than continuous.

Parts (b) and **(c)** both required the use of linear interpolation, but far more candidates were successful with **part (c)** than **part (b)**. There were a number of different, correct approaches used in both parts, but some candidates seemed to be trying to employ the sorts of techniques that would work in **part (c)** when they approached **part (b)**. Those that applied a correct technique in **part (b)** sometimes forgot to give their answer as a percentage or found the percentage of cars with a fuel consumption less, rather than more, than 9 *l*/100 km. Many fully correct solutions were seen to **part (c)**.

Most candidates used the information provided together with their answer to the previous part to draw an accurate box-and-whisker diagram in **part (d)**. A small number of candidates plotted the median exactly halfway between the lower and upper quartile, rather than at the position of their median from **part (c)**.

Most candidates were able to give at least one correct comparison of the fuel consumptions of the cars and the pickup trucks in **part (e)**. To score both marks consideration should be given to comparing both the central tendency and a measure of spread, giving that comparison in the context of the question. Some candidates, however, effectively gave the same comparison twice: for example, stating, firstly, that the fuel consumption of the cars is generally lower and then, secondly, that the fuel consumption of the pickup trucks is generally higher; or, firstly, comparing the two ranges and concluding that the variation in fuel consumption

is equal for the two types of vehicle and, secondly, comparing the two interquartile ranges and drawing the same conclusion.

Question 9

In **part (a),** quite a large number of candidates interpreted the price relative as telling them that there was no change in the 'expenditure' rather than that was no change in the 'cost'. This error has also been seen in previous years, where candidates assume that a price relative is giving information about expenditure rather than realising that it is independent of the amount used and is therefore only giving information about the cost. For full marks it was also necessary to explain that the figure related to electricity between 2016 and 2019.

In **part (b),** candidates were shown a calculation that did not make any use of weights, and were asked why it was not likely to produce a good estimate. It was surprisingly rare to see answers that referred to the lack of weights, with some candidates simply saying, 'the weights may have changed', an answer which appeared to assume that weights had in fact been used. It was extremely rare to see candidates also noting that the expenditures for each category (given at the start of the question), which could have been used for weights, were very different from each other and therefore their omission would have a large effect on the result.

Many candidates correctly calculated the price relative for electricity in 2020 using the unit costs provided in **part (c)**. It was quite common, however, to see the price relative for other costs given as 101 rather than 99.

Most candidates used a correct method to find the weighted aggregate cost index in **part (d)(i)**, although it was quite common to see this given to three significant figures rather than to the one decimal place requested in the question. A surprisingly common error seen in **part (d)(ii)** was for candidates to recalculate a mean price relative for 2020 using the method that they had been asked to criticise in **part (b)**. However, many fully correct solutions were also seen.

In **part (e)**, many candidates correctly identified *B* and *D* as the two explanations that could not be used to explain the inaccurate result, but only about half of these were able to provide the reason, namely that these figures had already been accounted for in the calculation.

Question 10

Many candidates in **part (a)** demonstrated some understanding of mutually exclusive and independent events, but it was rare to see a fully correct answer. The most commonly seen partially correct answer was *C*&*D* as a correct pair of mutually exclusive events and *A*&*B* as the correct pair of independent events. Missing often, therefore, was *A*&*D* and *B*&*C* from the list of mutually exclusive events.

Many candidates did not seem to be using their answers to **part (a)** when it came to answering **part (b)**. So, for example, it was common to see non-zero answers to **part (b)(i)**, even from candidates who had correctly identified *A*&*D* as a pair of mutually exclusive in **part (a)**. More candidates were successful with **part (b)(ii)** than the rest of **part (b)**, although some candidates did add rather than multiply the required probabilities. In **part (b)(iii)** many candidates correctly found P(*C*), but instead of simply adding this to P(*B*) they often also subtracted the product of the probabilities, even if they had identified *B*&*C* as a pair of mutually exclusive in **part (a)**.

Part (c) of this question was often well done, with those that did not score full marks often scoring partial credit for recognising, and demonstrating in their calculation, that selecting one white counter can happen in three ways and/or for recognising that the counters are not replaced, demonstrated by having $n(n-1)(n-2)$ in the denominator of their product.

Part (d) was a challenging question. Some candidates correctly identified that to achieve the given aim of each container having counters of only one colour, three black counters must be removed from the bag and then two white counters must be removed from the box. Some candidates were able to find the probability of removing three black counters from the bag correctly. A common error was to continue to assume that the total number of counters in the box had remained at 7 rather than that it had increased to 10. A further common error was for the probability of selecting three black counters from the bag to be added to, rather than multiplied by, the probability of selecting two white counters from the box. Some candidates did score full marks for this question, but also it was left blank in quite a large number of cases.