STATISTICS

Paper 4040/12
Paper 1

Key messages

Statistics is a subject dealing with real-life situations, and the nature of these situations should not be ignored when embarking on calculations required by the question.

After obtaining the result of a calculation the candidate should pause to consider whether or not the result seems possible and reasonable for the practical situation of the question.

In answering the writing parts of questions, which may involve explanations, it is important to focus on the specific situation of the question, and not simply reproduce something learned and hope it may be relevant.

The answer to any question where a probability has to be found cannot exceed 1.

General comments

The standard of work involving calculations of a routine nature was again generally good. This was particularly true of reading a pie chart (see **Question 3** below), finding the equation of a line of best fit (see **Question 7** below), finding crude and standardised rates (see **Question 8** below) and calculating the mean and standard deviation of an ungrouped frequency distribution (see **Question 10** below).

On some questions performance varied quite a lot (See **Questions 4** and **5** below). Very few candidates obtained full marks on all the probability elements on the paper (see **Questions 6** and **8(f)** below).

It has to be emphasised again, as it has been regularly in these reports, that Statistics is a subject which is applied to real-life situations, and is different from Pure Mathematics. Having obtained the answer to a question, candidates should think whether or not it is a reasonable one for the situation of the question. Having some appreciation of the relative sizes of numbers relevant to a particular situation is of help here: having some appreciation, for example, of how different in size something of height 5 centimetres or 5 metres or 50 metres will look. There were instances in this paper where some highly unrealistic answers were presented, answers which should have given the candidate at least pause for thought about the working they had done leading to these answers (see **Questions 7(c)** and **10(e)** below).

Comments on specific questions

Question 1

Responses to this question on basic sampling methods were mixed: there were many fully correct answers, but also many with one or more errors. ‘Systematic’ and ‘stratified’ were sometimes interchanged.

Question 2

Good knowledge was shown of how to find the mode and median of a small set of data, though finding the upper quartile proved to be more problematic. Overwhelmingly the main limitation in answers to this question was in responses to **part (b)**. Many candidates were able to offer perfectly correct *general* reasons why the mode is not a good measure of central tendency. But such answers did not address the question as to why it is not a good measure ‘in this case’. Few were able to say that it is not a good measure of central tendency simply because it is not central in the distribution, but at one end of it.

Question 3

Very good knowledge was shown on the interpretation of this pie chart, and on finding the radius of a comparative chart, even though the totals for the amounts of electricity generated by the two companies were not given. There were many fully correct answers.

Question 4

There were mixed answers to **part (a)**, as some candidates did not appear to know the form that a two-way table takes, and these earned none of the first four marks. Those who were able to form the table properly usually did so accurately, and values in all nine cells were entered accurately. There were many good answers to **part (b)**, with observations that cloud was most common in the morning, and sun most common in the evening, being seen regularly. The most serious limitation overall was in answers to **part (c)**. Most candidates understood that in losing the original data it became impossible to say what happened day to day, or throughout the course of particular days, but did not illustrate this with specific reference to the data. Only a few pointed to the fact, for example, that there were three consecutive days where it rained unchangingly throughout these days.

Question 5

Many fully correct histograms were seen in **part (a)**, but there were many also where all the column heights were drawn at frequency values. In the latter case only one of the five available marks could be awarded. Answers were mixed in **part (b)** also. As with any grouped frequency distribution it was necessary to work with class mid-points when estimating the total; yet a substantial number of candidates worked with class widths, or class limits.

Question 6

For success in this question it was essential to recognise at the outset that, if the probability of a customer failing to appear for their appointment is 0.1, then the probability of them appearing is 0.9.

Many candidates did in fact use these probabilities in **part (a)**, but in an incomplete or erroneous way. In **part (a)(i)** the product 0.9×0.1 for Cleo was seen often, but the 0.9^3 for Tony was either omitted totally, or added, instead of being multiplied with that for Cleo. A similar error was seen frequently in answers to **part (a)(ii)**, where the correct probabilities relating to the two hairdressers were incorrectly added. In **part (b)** it was commonly overlooked that, in order for a customer to accept a drink they had first to appear for their appointment, so 0.9 as well as 0.75 had to be used.

Overall the question was not well answered: some candidates did not use 0.9 or 0.1 at all; and others presented answers greater than 1. If the arithmetic in a probability question leads to an answer in excess of 1 the candidate should realise immediately that something must be wrong, and should look back over their work to try to find the error and correct it.

Question 7

Some very good work was seen on this question with candidates showing sound knowledge on how to find, and use, the equation of a line of best fit to experimental data. There were many fully correct answers in **parts (a)** to **(e)**. The exceptions usually occurred when errors were made in finding the gradient of the line, either through inversion of the correct expression, or using data points instead of the averages. Incorrect equations for the line of best fit sometimes meant impossible values were found in answers to **part (c)**. It is certainly not expected that candidates will have detailed knowledge of the heights of giraffes; but it *is* expected that they should have some appreciation of the sizes of numbers, so that it can be realised when an answer is unreasonable, and work can be checked to find an error. As examples which were seen, the height of a giraffe at birth is not going to be negative; nor is a giraffe ever going to grow to be 25 metres tall.

The last two **parts**, **(f)** and **(g)**, were less well answered. The key fact was that the plot of the data points revealed the relationship between age and height to be slightly non-linear.

Question 8

This was a new context for crude and standardised rates, but almost all candidates showed their command of the subject by applying their knowledge successfully to it. There were many fully correct answers to the first three parts of the question. Most candidates also knew the calculations to be done in **part (d)**, but many either only named the striker, and not the goals scored, or gave a non-integer value for the goals scored. In **part (e)**, for this context, it was a high value for the standardised rate that was most desirable.

The concluding probability part of the question proved problematic for most candidates. Much complicated arithmetic was seen for a situation that in fact required very little. If Alonso and Diame were to play in the first half of the match, then Benjani and Camara would automatically play in the second. Thus the problem reduced simply to the probability of choosing two particular players from four. Few candidates apparently recognised this.

Question 9

Throughout this question, those candidates who demonstrated their thinking by drawing lines on the graph are to be commended. This is good practice as it enables Examiners to award marks for method when the numerical value of an answer might be incorrect.

There were many good answers to **part (a)** on finding the required measures by reading a cumulative frequency graph accurately. Care is always needed, however, in such questions, over using the correct total frequency. A common error was working with the maximum value on the cumulative frequency axis (80), and not the total number of observations (76).

Part (b) was also reasonably well answered, though in **part (b)(ii)** the incorrect method of taking the mean of 4.0 and 6.0 was sometimes seen. Good understanding was also shown in answers to **part (c)(i)**, though less so in answers to **part (c)(ii)**. Here it was vital to understand that, in one method of solution, a reading from the graph at 2.5 mg/l rather than 5.5 mg/l had to be taken. The best answers to both parts of **(c)** concluded with the formal statement of an inequality, showing that further action was, or was not, necessary.

Several possible improvements were accepted in answers to **part (d)**. The most straightforward was that instead of taking just one measurement each day Hadiya should have taken several, and from them calculated a mean. Only a small number of such answers was seen.

Question 10

Parts (a) and **(b)** were answered very well by almost all candidates, with all necessary working laid out clearly. The main limitation, seen quite often, was in finding the standard deviation in **part (b)** to the required accuracy. The loss of accuracy by a small amount was the result of using only a three significant figure value for the mean in the standard deviation formula.

Part (c) was very much less well answered. In each part one choice had to be made and one reason had to be given. Where more than one of department and/or reason was given it was an indication to Examiners of limitations in the candidate's understanding.

The questions on the Venn diagram in **part (d)** were reasonably well answered, most errors occurring in **part (d)(ii)**. But many answers to the final question in **part (e)** showed much serious misunderstanding of what the question asked. The people in the question had registered in one, two or three of the named departments. Thus the mean number of departments in which they had registered had to be somewhere between one and three. Many candidates simply added the numbers in the diagram and divided by either 3 or 7, giving answers of 59.3 or 25.4. Here a little reflection on the impossibility of such answers for the situation of the question and what it asked should have caused the candidate to think again, to review their work, and to find and correct the error.

STATISTICS

Paper 4040/13
Paper 1

Key messages

Statistics is a subject dealing with real-life situations, and the nature of these situations should not be ignored when embarking on calculations required by the question.

After obtaining the result of a calculation the candidate should pause to consider whether or not the result seems possible and reasonable for the practical situation of the question.

In answering the writing parts of questions, which may involve explanations, it is important to focus on the specific situation of the question, and not simply reproduce something learned and hope it may be relevant.

The answer to any question where a probability has to be found cannot exceed 1.

General comments

The standard of work involving calculations of a routine nature was again generally good. This was particularly true of reading a pie chart (see **Question 3** below), finding the equation of a line of best fit (see **Question 7** below), finding crude and standardised rates (see **Question 8** below) and calculating the mean and standard deviation of an ungrouped frequency distribution (see **Question 10** below).

On some questions performance varied quite a lot (See **Questions 4** and **5** below). Very few candidates obtained full marks on all the probability elements on the paper (see **Questions 6** and **8(f)** below).

It has to be emphasised again, as it has been regularly in these reports, that Statistics is a subject which is applied to real-life situations, and is different from Pure Mathematics. Having obtained the answer to a question, candidates should think whether or not it is a reasonable one for the situation of the question. Having some appreciation of the relative sizes of numbers relevant to a particular situation is of help here: having some appreciation, for example, of how different in size something of height 5 centimetres or 5 metres or 50 metres will look. There were instances in this paper where some highly unrealistic answers were presented, answers which should have given the candidate at least pause for thought about the working they had done leading to these answers (see **Questions 7(c)** and **10(e)** below).

Comments on specific questions

Question 1

Responses to this question on basic sampling methods were mixed: there were many fully correct answers, but also many with one or more errors. ‘Systematic’ and ‘stratified’ were sometimes interchanged.

Question 2

Good knowledge was shown of how to find the mode and median of a small set of data, though finding the upper quartile proved to be more problematic. Overwhelmingly the main limitation in answers to this question was in responses to **part (b)**. Many candidates were able to offer perfectly correct *general* reasons why the mode is not a good measure of central tendency. But such answers did not address the question as to why it is not a good measure ‘in this case’. Few were able to say that it is not a good measure of central tendency simply because it is not central in the distribution, but at one end of it.

Question 3

Very good knowledge was shown on the interpretation of this pie chart, and on finding the radius of a comparative chart, even though the totals for the amounts of electricity generated by the two companies were not given. There were many fully correct answers.

Question 4

There were mixed answers to **part (a)**, as some candidates did not appear to know the form that a two-way table takes, and these earned none of the first four marks. Those who were able to form the table properly usually did so accurately, and values in all nine cells were entered accurately. There were many good answers to **part (b)**, with observations that cloud was most common in the morning, and sun most common in the evening, being seen regularly. The most serious limitation overall was in answers to **part (c)**. Most candidates understood that in losing the original data it became impossible to say what happened day to day, or throughout the course of particular days, but did not illustrate this with specific reference to the data. Only a few pointed to the fact, for example, that there were three consecutive days where it rained unchangingly throughout these days.

Question 5

Many fully correct histograms were seen in **part (a)**, but there were many also where all the column heights were drawn at frequency values. In the latter case only one of the five available marks could be awarded. Answers were mixed in **part (b)** also. As with any grouped frequency distribution it was necessary to work with class mid-points when estimating the total; yet a substantial number of candidates worked with class widths, or class limits.

Question 6

For success in this question it was essential to recognise at the outset that, if the probability of a customer failing to appear for their appointment is 0.1, then the probability of them appearing is 0.9.

Many candidates did in fact use these probabilities in **part (a)**, but in an incomplete or erroneous way. In **part (a)(i)** the product 0.9×0.1 for Cleo was seen often, but the 0.9^3 for Tony was either omitted totally, or added, instead of being multiplied with that for Cleo. A similar error was seen frequently in answers to **part (a)(ii)**, where the correct probabilities relating to the two hairdressers were incorrectly added. In **part (b)** it was commonly overlooked that, in order for a customer to accept a drink they had first to appear for their appointment, so 0.9 as well as 0.75 had to be used.

Overall the question was not well answered: some candidates did not use 0.9 or 0.1 at all; and others presented answers greater than 1. If the arithmetic in a probability question leads to an answer in excess of 1 the candidate should realise immediately that something must be wrong, and should look back over their work to try to find the error and correct it.

Question 7

Some very good work was seen on this question with candidates showing sound knowledge on how to find, and use, the equation of a line of best fit to experimental data. There were many fully correct answers in **parts (a)** to **(e)**. The exceptions usually occurred when errors were made in finding the gradient of the line, either through inversion of the correct expression, or using data points instead of the averages. Incorrect equations for the line of best fit sometimes meant impossible values were found in answers to **part (c)**. It is certainly not expected that candidates will have detailed knowledge of the heights of giraffes; but it *is* expected that they should have some appreciation of the sizes of numbers, so that it can be realised when an answer is unreasonable, and work can be checked to find an error. As examples which were seen, the height of a giraffe at birth is not going to be negative; nor is a giraffe ever going to grow to be 25 metres tall.

The last two **parts**, **(f)** and **(g)**, were less well answered. The key fact was that the plot of the data points revealed the relationship between age and height to be slightly non-linear.

Question 8

This was a new context for crude and standardised rates, but almost all candidates showed their command of the subject by applying their knowledge successfully to it. There were many fully correct answers to the first three parts of the question. Most candidates also knew the calculations to be done in **part (d)**, but many either only named the striker, and not the goals scored, or gave a non-integer value for the goals scored. In **part (e)**, for this context, it was a high value for the standardised rate that was most desirable.

The concluding probability part of the question proved problematic for most candidates. Much complicated arithmetic was seen for a situation that in fact required very little. If Alonso and Diame were to play in the first half of the match, then Benjani and Camara would automatically play in the second. Thus the problem reduced simply to the probability of choosing two particular players from four. Few candidates apparently recognised this.

Question 9

Throughout this question, those candidates who demonstrated their thinking by drawing lines on the graph are to be commended. This is good practice as it enables Examiners to award marks for method when the numerical value of an answer might be incorrect.

There were many good answers to **part (a)** on finding the required measures by reading a cumulative frequency graph accurately. Care is always needed, however, in such questions, over using the correct total frequency. A common error was working with the maximum value on the cumulative frequency axis (80), and not the total number of observations (76).

Part (b) was also reasonably well answered, though in **part (b)(ii)** the incorrect method of taking the mean of 4.0 and 6.0 was sometimes seen. Good understanding was also shown in answers to **part (c)(i)**, though less so in answers to **part (c)(ii)**. Here it was vital to understand that, in one method of solution, a reading from the graph at 2.5 mg/l rather than 5.5 mg/l had to be taken. The best answers to both parts of **(c)** concluded with the formal statement of an inequality, showing that further action was, or was not, necessary.

Several possible improvements were accepted in answers to **part (d)**. The most straightforward was that instead of taking just one measurement each day Hadiya should have taken several, and from them calculated a mean. Only a small number of such answers was seen.

Question 10

Parts (a) and **(b)** were answered very well by almost all candidates, with all necessary working laid out clearly. The main limitation, seen quite often, was in finding the standard deviation in **part (b)** to the required accuracy. The loss of accuracy by a small amount was the result of using only a three significant figure value for the mean in the standard deviation formula.

Part (c) was very much less well answered. In each part one choice had to be made and one reason had to be given. Where more than one of department and/or reason was given it was an indication to Examiners of limitations in the candidate's understanding.

The questions on the Venn diagram in **part (d)** were reasonably well answered, most errors occurring in **part (d)(ii)**. But many answers to the final question in **part (e)** showed much serious misunderstanding of what the question asked. The people in the question had registered in one, two or three of the named departments. Thus the mean number of departments in which they had registered had to be somewhere between one and three. Many candidates simply added the numbers in the diagram and divided by either 3 or 7, giving answers of 59.3 or 25.4. Here a little reflection on the impossibility of such answers for the situation of the question and what it asked should have caused the candidate to think again, to review their work, and to find and correct the error.

STATISTICS

Paper 4040/22
Paper 2

Key messages

To be successful in this examination candidates need to read the questions carefully and provide clearly set out solutions, particularly in multistage problems. Final answers should always be checked to make sure that they appear to be sensible. For example, the answer to a probability question is never going to be greater than 1. Clear working is also essential in questions where an answer is given. Diagrams should always be labelled, including with any units. Interpretation of diagrams or calculations should always be done in the context of the problem described.

General comments

Some well-structured responses were seen, particularly in **Question 7(b)**. Sometimes essential working was missing from some candidates' work; this was particularly important in **Question 8(a)**, where the answer was given.

Some marks appeared to be lost on this paper due to questions not being read sufficiently carefully. Examples where this sometimes occurred were in **Questions 2(a)**, **3(a)**, **7(f)**, **8(d)** and **10(b)(i)**.

Most candidates used appropriate scales in questions requiring the drawing of a diagram and made accurate plots. As in previous series, the labelling of axes was sometimes missing or incomplete.

There was an improvement, compared to previous series, in the interpretation in context of diagrams, most notably in **Question 4(a)**.

Comments on specific questions

Question 1

Question 1 was a standard question on scaling data to a given mean and standard deviation. Those that set the work out by equating standardised marks, so by solving $(x – 50)/12 = (41 – 60.5)/7.8$ in **part (a)**, for example, were usually successful. In **part (b)** a small number of candidates produced a correct equation in the unknown standard deviation, but made errors solving for an unknown in the denominator. Some candidates were unable to produce an equation with the unknown appearing twice, as was required in **part (c)** and some candidates omitted this part of the question.

Question 2

Most candidates were successful with **part (a)**, but a surprising number gave an incorrect answer of ½, perhaps coming from doing $P(B')$ rather than $P(A')$.

Part (b) was answered well by most candidates, many of whom went on to answer **part (c)** correctly. Some candidates omitted to subtract $P(A\cap B)$ in **part (c)**, leaving them with a probability greater than 1; this should have alerted them to the fact that they had made an error.

In **part (d)**, those that knew that the pair of mutually exclusive events were A and A' were usually able to express the reason correctly, although a few gave the reason as $P(A) + P(A') = 1$. Many candidates incorrectly gave just one of the events as their answer, rather than picking out a pair of events.

Question 3

It was pleasing to see some good attempts at **part (a)**, with many candidates realising the need to subtract the two given item numbers. Some candidates went on to find the smallest item number rather than the smallest possible size of the sample, so careful reading of the question was important here.

Most candidates were able to correctly calculate the sample sizes in **part (b)(i)**. In **part (b)(ii)**, some candidates found a simple random sample rather than a sample stratified by machine and a few candidates repeated the second 04 in their sample. However, most answered this part correctly.

Question 4

It was pleasing to see some good answers to **part (a)** compared to similar questions in the past. Many candidates correctly identified that the paw lengths of the coyotes were generally longer, giving their comparison in the context of the question. It was less common to see a correct comparison of the spread, namely that the paw lengths of the coyotes were more varied. Instead, many candidates attempted to make a comparison of the frequencies, incorrectly stating that there were more red foxes.

Parts (b) and **(c)** proved to be much more challenging. Some candidates were able to explain that frequency polygons can be displayed together allowing for easy comparison, but many gave incorrect answers, such as the original data is not lost. Others suggested that frequency polygons were easier to construct or were more accurate, which did not gain any credit. Advantages of a particular method of display should always be given in terms of advantages for interpretation rather than ease of drawing.

The final part proved to be very challenging, with most candidates thinking that there were some red foxes in the sample with a paw length of 7.1 cm. Candidates needed to realise that frequency polygon plots are at midpoints of class intervals and therefore the plot at (7.5, 0) implied that no red foxes had paw lengths greater than 7 cm.

Question 5

Some candidates correctly identified, in **part (a)**, that figure 1 was misleading because the vertical axis did not start at zero. Some went further and explained that the impact of this was to make it look as if the number of employees had doubled over the ten-year period. It was less common to see the misleading aspect of figure 2 identified correctly, namely that the scale showing the annual salaries was not a linear scale; indeed, some candidates used a similarly non-linear scale in **part (c)** when drawing their own diagram for 2012. Some candidates did however point out that this non-linear scale gave the incorrect impression that the data was symmetrical.

Most candidates correctly found the median in **part (b)** and made a good attempt at the box-and-whisker diagram, using a linear scale in **part (c)**. The most common error was for the horizontal axis to not be fully labelled. The most common missing feature was 'thousands', which was required if the scale had been labelled as 7, 8, 9 etc. It was also necessary to label the axis as representing 'salary' with the units given as dollars. Other errors seen were for the scale to be labelled as 70, 80, 90 etc., with no mention of 'hundreds' and sometimes a scale of this sort which started with 70, 80 and 90 continued with 10, 11 and 12. Most of these candidates had correctly applied the key when they found the median in **part (b)**, but did not continue with applying the key correctly when illustrating the data in the box-and-whisker diagram. Those that had an appropriate linear scale usually found the quartiles, highest and lowest values correctly.

Question 6

The correct total number of springs was usually seen in **part (a)**. This value, however, was frequently not used in the solution in **part (b)**. Many candidates earned at least one of the available marks in **part (b)**, usually by demonstrating that they understood that the spring selection took place without replacement, but it was often not 44×43 that was seen in the denominator. Fully correct solutions were seen only from the strongest responses.

Question 7

Many fully correct descriptions of the type of data were seen in **part (a)**. **Part (b)** required the use of linear interpolation to find the interquartile range. Most candidates earned all or most of the marks in this question. A few found the upper quartile and forgot to continue to find the interquartile range. In a multistage problem of this sort, it is important to check at the end of the solution that all parts of the question have been answered. Some only earned the marks for finding the position of the upper quartile and/or for subtracting the lower quartile from their upper quartile. For those that did not get the answer fully correct, it was usually easy for the Examiner to follow the candidate's work and award marks for the correct parts of the method.

Part (c) of this question proved to be another very challenging question. It was necessary to look both at the frequencies in the table, and the intervals to which they relate. The class interval $20 \leqslant t < 40$ is a wider class than the others and therefore the 13 values in this class are likely to include some extreme (low) values of times to complete the challenge. The mean is therefore likely to be less than the median. It was necessary for reasons to reference that the extreme times were low.

Parts (d) and **(f)** required further use of linear interpolation, with each part getting progressively more difficult. Most candidates were successful with **part (d)**, but **part (f)** proved much more challenging. Those that did not score full marks were sometimes able to score partial credit for taking correct account of the 2½ minutes to read the instructions. Some gave, as the final answer, the number who now qualify, rather than how many more now qualify. Again, careful rereading of the question would be helpful.

In **part (e)** candidates needed to understand the effect on both the median and the interquartile range of the 2½ minute adjustment, and apply that change only to the median. The most common error was for 2½ minutes to be subtracted from the interquartile range as well as the median. Some candidates did not provide attempts at **parts (d)**, **(e)** or **(f)** of this question.

Question 8

In **part (a)** candidates were instructed to show a given result. It was therefore important that clear calculations leading to the values 960 and 2160 were shown. Many did show the calculations clearly, but some candidates did not provide any working for those values and others showed insufficient/unclear working. The expected working for the value 960, for example, was $40 \times 24\,000 \div 1000$.

In **part (b)** most candidates used an appropriate linear scale and correctly drew the sectional bar chart. As with **Question 5(b)**, the most common error was for the labelling of the axis to be missing. In this case the label 'percentage' or 'per cent' was often omitted.

Part (c) was very challenging, as candidates needed to use their sectional bar chart to see that the weight for fuel was over 50 per cent and then also explain that the decrease in fuel (8 per cent) was greater than the two increases (7 per cent and 2 per cent). Some candidates provided a partial explanation by saying that the weight for fuel was over 50 per cent, but many incorrectly thought that the overall costs would increase.

Most candidates correctly found the price relatives from the given information in **part (d)**, with the most common error being an increase of 16 per cent to 116 for tax and insurance in 2022, rather than a reduction of 16 per cent. Reading the question carefully to check whether each value represents an increase, or a decrease, is important. Most candidates used a correct method to find the weighted cost of driving index in **part (e)**.

Many correct answers were seen in **part (f)**, usually relating to a change in the distance travelled or the fuel consumption of the new car. Some did, however, incorrectly give a change in price as the explanation.

Question 9

This was a more challenging question than usual on moving averages, so it was pleasing to see some very good attempts.

Part (a) was straightforward, with the most common error being that the line segments joining the plots were sometimes missing or sometimes did not join the plots consecutively. Most of the plots themselves were accurately drawn. Some candidates did convey that the trend should come from considering all the data rather than simply the last two readings in **part (b)**, but others talked about other specific readings.

It was very pleasing to see so many fully correct solutions in **part (c)**, as this was a much less structured question than many on this topic have been in the past. Firstly, candidates needed to realise that they must find moving average values. Most candidates did so, with some credit given to those that chose an incorrect n-point moving average. To find the correct value for n, namely 3, it was necessary to look at the pattern in the plotted data in **part (a)**; this was a much easier task for those that had joined the plots correctly with straight line segments. A small number only found four 3-point moving average values, but many correctly found all ten values and plotted them correctly with an appropriate trend line.

Again, **part (d)** was more challenging than similar questions in the past. It was necessary to decide which of the given times corresponded to the same season as time 13. Some candidates correctly found differences between readings and moving average values but did so for all the times rather than just the appropriate times of 4, 7 and 10. Some credit was available for these candidates if they used their seasonal component correctly with their trend line reading, as many did. Fully correct solutions were seen from the strongest responses.

Question 10

Many candidates found question 10 challenging, and some did not complete the final parts of the question.

The most common incorrect solution seen in **part (a)(i)** was $1 - (\frac{1}{4})^3$ rather than $(1 - \frac{1}{4})^3$. Those that got **part (i)** correct usually went on to get **part (a)(ii)** correct. Many did not realise that the answer to **part (b)(i)** came from summing the two previous answers, with many just giving the answer to **part (a)(ii)** here. Careful reading of both the information above the table and the information in the table was required. Candidates were more successful with **parts (b)(ii)** and **(b)(iii)**, although some weaker scripts contained answers greater than 1 for these probabilities.

It was from **part (c)** onwards that some candidates left the answer spaces blank. The most commonly seen error was for the values in the table to be the numbers on the final squares, rather than the prize values. Those that gave prize values often realised that the probabilities in their table needed to sum to 1, and some fully correct solutions were seen.

Those that had prize values in their table in **(c)** were often able to use the correct method in **part (d)**, realising that to be a fair game the amount charged should be equal to the expected prize. These candidates were also often able to form an appropriate equation in **part (e)**, using a letter to represent the unknown prize.

Part (f) was challenging, so it was pleasing to see some correct explanations, including from candidates that had made earlier errors in this question. The most common incorrect answers seen were explanations such as ‘the amount charged to play has doubled’ without reference to the very small probability of finishing on square 8.

STATISTICS

Paper 4040/22
Paper 2

Key messages

To be successful in this examination candidates need to read the questions carefully and provide clearly set out solutions, particularly in multistage problems. Final answers should always be checked to make sure that they appear to be sensible. For example, the answer to a probability question is never going to be greater than 1. Clear working is also essential in questions where an answer is given. Diagrams should always be labelled, including with any units. Interpretation of diagrams or calculations should always be done in the context of the problem described.

General comments

Some well-structured responses were seen, particularly in **Question 7(b)**. Sometimes essential working was missing from some candidates' work; this was particularly important in **Question 8(a)**, where the answer was given.

Some marks appeared to be lost on this paper due to questions not being read sufficiently carefully. Examples where this sometimes occurred were in **Questions 2(a)**, **3(a)**, **7(f)**, **8(d)** and **10(b)(i)**.

Most candidates used appropriate scales in questions requiring the drawing of a diagram and made accurate plots. As in previous series, the labelling of axes was sometimes missing or incomplete.

There was an improvement, compared to previous series, in the interpretation in context of diagrams, most notably in **Question 4(a)**.

Comments on specific questions

Question 1

Question 1 was a standard question on scaling data to a given mean and standard deviation. Those that set the work out by equating standardised marks, so by solving $(x – 50)/12 = (41 – 60.5)/7.8$ in **part (a)**, for example, were usually successful. In **part (b)** a small number of candidates produced a correct equation in the unknown standard deviation, but made errors solving for an unknown in the denominator. Some candidates were unable to produce an equation with the unknown appearing twice, as was required in **part (c)** and some candidates omitted this part of the question.

Question 2

Most candidates were successful with **part (a)**, but a surprising number gave an incorrect answer of ½, perhaps coming from doing $P(B')$ rather than $P(A')$.

Part (b) was answered well by most candidates, many of whom went on to answer **part (c)** correctly. Some candidates omitted to subtract $P(A \cap B)$ in **part (c)**, leaving them with a probability greater than 1; this should have alerted them to the fact that they had made an error.

In **part (d)**, those that knew that the pair of mutually exclusive events were A and A' were usually able to express the reason correctly, although a few gave the reason as $P(A) + P(A') = 1$. Many candidates incorrectly gave just one of the events as their answer, rather than picking out a pair of events.

Question 3

It was pleasing to see some good attempts at **part (a)**, with many candidates realising the need to subtract the two given item numbers. Some candidates went on to find the smallest item number rather than the smallest possible size of the sample, so careful reading of the question was important here.

Most candidates were able to correctly calculate the sample sizes in **part (b)(i)**. In **part (b)(ii)**, some candidates found a simple random sample rather than a sample stratified by machine and a few candidates repeated the second 04 in their sample. However, most answered this part correctly.

Question 4

It was pleasing to see some good answers to **part (a)** compared to similar questions in the past. Many candidates correctly identified that the paw lengths of the coyotes were generally longer, giving their comparison in the context of the question. It was less common to see a correct comparison of the spread, namely that the paw lengths of the coyotes were more varied. Instead, many candidates attempted to make a comparison of the frequencies, incorrectly stating that there were more red foxes.

Parts (b) and **(c)** proved to be much more challenging. Some candidates were able to explain that frequency polygons can be displayed together allowing for easy comparison, but many gave incorrect answers, such as the original data is not lost. Others suggested that frequency polygons were easier to construct or were more accurate, which did not gain any credit. Advantages of a particular method of display should always be given in terms of advantages for interpretation rather than ease of drawing.

The final part proved to be very challenging, with most candidates thinking that there were some red foxes in the sample with a paw length of 7.1 cm. Candidates needed to realise that frequency polygon plots are at midpoints of class intervals and therefore the plot at (7.5, 0) implied that no red foxes had paw lengths greater than 7 cm.

Question 5

Some candidates correctly identified, in **part (a)**, that figure 1 was misleading because the vertical axis did not start at zero. Some went further and explained that the impact of this was to make it look as if the number of employees had doubled over the ten-year period. It was less common to see the misleading aspect of figure 2 identified correctly, namely that the scale showing the annual salaries was not a linear scale; indeed, some candidates used a similarly non-linear scale in **part (c)** when drawing their own diagram for 2012. Some candidates did however point out that this non-linear scale gave the incorrect impression that the data was symmetrical.

Most candidates correctly found the median in **part (b)** and made a good attempt at the box-and-whisker diagram, using a linear scale in **part (c)**. The most common error was for the horizontal axis to not be fully labelled. The most common missing feature was ‘thousands’, which was required if the scale had been labelled as 7, 8, 9 etc. It was also necessary to label the axis as representing ‘salary’ with the units given as dollars. Other errors seen were for the scale to be labelled as 70, 80, 90 etc., with no mention of ‘hundreds’ and sometimes a scale of this sort which started with 70, 80 and 90 continued with 10, 11 and 12. Most of these candidates had correctly applied the key when they found the median in **part (b)**, but did not continue with applying the key correctly when illustrating the data in the box-and-whisker diagram. Those that had an appropriate linear scale usually found the quartiles, highest and lowest values correctly.

Question 6

The correct total number of springs was usually seen in **part (a)**. This value, however, was frequently not used in the solution in **part (b)**. Many candidates earned at least one of the available marks in **part (b)**, usually by demonstrating that they understood that the spring selection took place without replacement, but it was often not 44 × 43 that was seen in the denominator. Fully correct solutions were seen only from the strongest responses.

Question 7

Many fully correct descriptions of the type of data were seen in **part (a)**. **Part (b)** required the use of linear interpolation to find the interquartile range. Most candidates earned all or most of the marks in this question. A few found the upper quartile and forgot to continue to find the interquartile range. In a multistage problem of this sort, it is important to check at the end of the solution that all parts of the question have been answered. Some only earned the marks for finding the position of the upper quartile and/or for subtracting the lower quartile from their upper quartile. For those that did not get the answer fully correct, it was usually easy for the Examiner to follow the candidate's work and award marks for the correct parts of the method.

Part (c) of this question proved to be another very challenging question. It was necessary to look both at the frequencies in the table, and the intervals to which they relate. The class interval $20 \leqslant t < 40$ is a wider class than the others and therefore the 13 values in this class are likely to include some extreme (low) values of times to complete the challenge. The mean is therefore likely to be less than the median. It was necessary for reasons to reference that the extreme times were low.

Parts (d) and **(f)** required further use of linear interpolation, with each part getting progressively more difficult. Most candidates were successful with **part (d)**, but **part (f)** proved much more challenging. Those that did not score full marks were sometimes able to score partial credit for taking correct account of the 2½ minutes to read the instructions. Some gave, as the final answer, the number who now qualify, rather than how many more now qualify. Again, careful rereading of the question would be helpful.

In **part (e)** candidates needed to understand the effect on both the median and the interquartile range of the 2½ minute adjustment, and apply that change only to the median. The most common error was for 2½ minutes to be subtracted from the interquartile range as well as the median. Some candidates did not provide attempts at **parts (d)**, **(e)** or **(f)** of this question.

Question 8

In **part (a)** candidates were instructed to show a given result. It was therefore important that clear calculations leading to the values 960 and 2160 were shown. Many did show the calculations clearly, but some candidates did not provide any working for those values and others showed insufficient/unclear working. The expected working for the value 960, for example, was $40 \times 24\,000 \div 1000$.

In **part (b)** most candidates used an appropriate linear scale and correctly drew the sectional bar chart. As with **Question 5(b)**, the most common error was for the labelling of the axis to be missing. In this case the label 'percentage' or 'per cent' was often omitted.

Part (c) was very challenging, as candidates needed to use their sectional bar chart to see that the weight for fuel was over 50 per cent and then also explain that the decrease in fuel (8 per cent) was greater than the two increases (7 per cent and 2 per cent). Some candidates provided a partial explanation by saying that the weight for fuel was over 50 per cent, but many incorrectly thought that the overall costs would increase.

Most candidates correctly found the price relatives from the given information in **part (d)**, with the most common error being an increase of 16 per cent to 116 for tax and insurance in 2022, rather than a reduction of 16 per cent. Reading the question carefully to check whether each value represents an increase, or a decrease, is important. Most candidates used a correct method to find the weighted cost of driving index in **part (e)**.

Many correct answers were seen in **part (f)**, usually relating to a change in the distance travelled or the fuel consumption of the new car. Some did, however, incorrectly give a change in price as the explanation.

Question 9

This was a more challenging question than usual on moving averages, so it was pleasing to see some very good attempts.

Part (a) was straightforward, with the most common error being that the line segments joining the plots were sometimes missing or sometimes did not join the plots consecutively. Most of the plots themselves were accurately drawn. Some candidates did convey that the trend should come from considering all the data rather than simply the last two readings in **part (b)**, but others talked about other specific readings.

It was very pleasing to see so many fully correct solutions in **part (c)**, as this was a much less structured question than many on this topic have been in the past. Firstly, candidates needed to realise that they must find moving average values. Most candidates did so, with some credit given to those that chose an incorrect n-point moving average. To find the correct value for n, namely 3, it was necessary to look at the pattern in the plotted data in **part (a)**; this was a much easier task for those that had joined the plots correctly with straight line segments. A small number only found four 3-point moving average values, but many correctly found all ten values and plotted them correctly with an appropriate trend line.

Again, **part (d)** was more challenging than similar questions in the past. It was necessary to decide which of the given times corresponded to the same season as time 13. Some candidates correctly found differences between readings and moving average values but did so for all the times rather than just the appropriate times of 4, 7 and 10. Some credit was available for these candidates if they used their seasonal component correctly with their trend line reading, as many did. Fully correct solutions were seen from the strongest responses.

Question 10

Many candidates found question 10 challenging, and some did not complete the final parts of the question.

The most common incorrect solution seen in **part (a)(i)** was $1 - (\frac{1}{4})^3$ rather than $(1 - \frac{1}{4})^3$. Those that got **part (i)** correct usually went on to get **part (a)(ii)** correct. Many did not realise that the answer to **part (b)(i)** came from summing the two previous answers, with many just giving the answer to **part (a)(ii)** here. Careful reading of both the information above the table and the information in the table was required. Candidates were more successful with **parts (b)(ii)** and **(b)(iii)**, although some weaker scripts contained answers greater than 1 for these probabilities.

It was from **part (c)** onwards that some candidates left the answer spaces blank. The most commonly seen error was for the values in the table to be the numbers on the final squares, rather than the prize values. Those that gave prize values often realised that the probabilities in their table needed to sum to 1, and some fully correct solutions were seen.

Those that had prize values in their table in **(c)** were often able to use the correct method in **part (d)**, realising that to be a fair game the amount charged should be equal to the expected prize. These candidates were also often able to form an appropriate equation in **part (e)**, using a letter to represent the unknown prize.

Part (f) was challenging, so it was pleasing to see some correct explanations, including from candidates that had made earlier errors in this question. The most common incorrect answers seen were explanations such as ‘the amount charged to play has doubled’ without reference to the very small probability of finishing on square 8.