

STATISTICS

Paper 4040/12
Paper 1

Key messages

Understanding the purpose of calculating a statistical measure is as important as actually being able to calculate it.

After obtaining the answer to a question the candidate should consider whether or not the result is reasonable for the practical situation of the question.

In answering the writing part of a question the candidate should understand properly the meaning of anything they may have learned before using it, and where necessary adapt it to the specific context and requirements of the question.

General comments

The quality of work involving routine numerical calculations was again generally high. This was particularly true of finding crude and standardised rates (see **Question 8** below), calculating the mean and standard deviation of a grouped frequency distribution (see **Question 9** below), and finding the equation of a line of best fit to experimental data (see **Question 11** below). Performance on the short-answer questions was also very good (see **Questions 1, 2 and 3** below).

Amongst the shorter questions it was rare to see full mark answers to the probability question (see **Question 6** below).

Candidates struggled with the writing and 'explain' questions. Often in their attempts at these questions responses were so vague as to have little meaning. The impression was gained that short phrases had been memorised, without full understanding of their meaning, to be produced in the hope of being what was required (see **Questions 8(d), 9(a) and 10(d)** below).

It needs to be emphasised again, as it has been regularly in these reports, that Statistics is a subject with context, used to analyse real-life situations; it is not just about manipulating abstract numbers. Candidates should think whether or not an answer is reasonable for a given situation. There were instances in this paper where unrealistic answers were presented, answers which should have given the candidate at least pause for thought (see **Questions 10(c) and 11(d)** below).

Comments on specific questions

Question 1

The pictogram was interpreted accurately by almost all candidates. The occasional error occurred in part **(b)**, where the answer given was for the number of women who were satisfied, rather than the difference between the number who were dissatisfied and this value.

Question 2

This also was answered well, with only an occasional error occurring in **part (a)(ii)**, resulting from the data not having been arranged in numerical order before identifying the two middle values.

Question 3

Almost all candidates were able to interpret the two-way table correctly. Where an error occurred it was usually in **part (e)**, with the denominator of the probability fraction being given as 40 rather than 14.

Question 4

Good answers to **part (a)** made a simple statement like 'as one variable increases the other variable also increases'. Nothing more than this was required. Answers which attempted to describe the pattern of plots on a scatter diagram showing positive correlation were often unclear and rarely satisfactory. It was also not enough to state only that the variables are directly proportional to each other. Whilst it is true that directly proportionality implies positive correlation, the converse is not the case.

Many completely correct answers to **part (b)** were seen. Where they were not it was usually because of the words 'weak' and 'strong' being omitted.

Question 5

Answers to this question showed most limitations in responses to **parts (a)(iii)** and **(a)(iv)**. Presumably because 'green', 'small' and 'special' were not seen on the diagram, many answers of zero were seen here. Such candidates had not been able, for example, to equate 'green' with not black. Candidates who were able to make such a deduction, there being only two categories for each of colour, size and type, almost always had fully correct answers.

Question 6

Correct deductions in **part (a)** were seen fairly often. Working showing the three possibilities for one cheese and two tomato, with products of three without replacement probabilities, produced the given answer. Incorrect contrived calculations resulting in the value $10/21$ were not only not credited, but acted to the disadvantage of the candidate: it was consequently not possible to extend the correct work which should have been done here to the other possible selections in **part (b)**.

Fully correct answers to **part (b)** were quite rare, though two or three marks were sometimes obtained by candidates who had been successful in **part (a)**. The main reason for this was that the possibility of selecting no cheese sandwiches was frequently overlooked.

Question 7

There was an overall improvement this time in the interpretation of histograms. Most candidates recognised that group frequencies could not necessarily be found by simply reading off the heights of the different bars: the labelling of the vertical axis was crucial in finding these frequencies. So even when answers to **parts (a)** and **(b)** were incorrect, some credit was given provided the values presented were in the right ratio (for example, if all the values presented were twice the correct values).

There were very few satisfactory answers to **part (c)**. Whilst many candidates could point out that a single bar only was used in each of the parts of **part (a)**, and that two bars were used in **part (b)**, this was not considered enough. The essential point was that frequencies for the complete single bars in **part (a)** were known, whilst the frequencies for the *parts* of the two bars in **part (b)** were unknown.

Question 8

Candidates continue to perform well in the routine calculations associated with crude and standardised rates, and full marks were frequently obtained in the first three parts of this fertility rates question. Given the facility with such calculations however, it was regrettable that not all were able to explain in **part (d)** the actual purpose of finding a standardised rate. A few candidates, not giving proper attention to the context, and presumably memorising work on past problems, referred to the health of towns and deaths.

In **part (e)** candidates had to select the appropriate rates to perform the necessary calculations, and performed exceptionally well in almost universally choosing the crude over the standardised. It has however to be pointed out that the question asked for the number of births; a good number of answers entered only the town in the answer line.

Question 9

The main problem with many answers to **part (a)** was that they tended to be so vague and general as to have little meaning. Responses which referred only to making things easier, quicker or being 'time-consuming' were not accepted. One of the best responses for a disadvantage, concise and clear, which many candidates produced, was 'the original data is lost'.

In contrast, answers to **parts (b)** and **(c)** were very good. In **part (c)** working was set out clearly leading to correct answers. When computing the standard deviation from this grouped frequency distribution many more candidates than has sometimes been the case in the past used the formula with Σfx^2 rather than that with $\Sigma f(x - \bar{x})^2$. From this they derived the benefit of not having to work with numbers having four decimal places.

Answers to **parts (d)** and **(e)** were more varied in quality. Candidates who realised in **part (d)** that the two different charges comprising the electricity bill had to be multiplied by different values, one by Σfx and the other by Σf , usually earned full marks. Most errors occurred in finding the charge due to electricity consumption: the \$0.18 was sometimes multiplied by 50, or the sum of the class mid-points, or even the class widths. An incorrect first step that was seen quite often was to add the \$0.18 and \$0.25 before carrying out any multiplication. A correct answer to **part (d)** was almost always followed by a correct answer to **part (e)**.

Question 10

At the outset it needs to be pointed out, as it has been for similar questions in the past, that care has to be taken in the use of the correct total frequency. Although it did not happen often, there were candidates who worked with the maximum value on the cumulative frequency axis (80), and not the total number of deliveries (72). This was in spite of the fact that the latter value could be seen on the graph and was also stated in the question.

Throughout the question those candidates who demonstrated their thinking by drawing lines on the graph are to be commended. Many did this and it enabled examiners to consider awarding marks for method where the numerical value of an answer might be incorrect.

There was a good number of accurate answers to **parts (a)** and **(b)**. The least well answered in these early questions was **part (a)(iv)**: in some answers the value 123 was incorporated into the calculation; in others the cumulative frequency corresponding to 123 km/h was found, the essential first step, but nothing more was done.

It was in **part (c)** where even those candidates with full marks to this point often fell into error. Instead of reading the cumulative frequency corresponding to 130 km/h it was read for 120 km/h. This resulted in an incorrect answer of about 55% being given. At this point, a little reflection on the nature of the situation should have alerted the candidate to the fact that this value was totally unrealistic. In **part (b)(i)** it was established that 29% of Buraid's deliveries were classified as 'fast', so now, after a reduction in their speed, that percentage had to be smaller, not larger.

In **part (d)**, most candidates were able to say that the interquartile range would be unchanged, but explanations as to why this was so were often inadequate. Short phrases of a very general nature, not related to the context, were often given, such as 'it is not affected by addition and subtraction'. Apart from the fact that there was no 'addition' (= increase in delivery speed here), such an answer made no reference, for example, to how the quartiles would change with deliveries 5 km/h slower, or how the amount of dispersion in the distribution would change with all data values 5 km/h smaller.

Question 11

In some cases the graphical work in **parts (a)** and **(b)** could have been improved by making the data point plotting clearer, and the averages plotted in a different way from the data points (for example, using crosses then circled points). But overall, very good work was seen in the first three parts of this question, with candidates showing sound knowledge on how to find the equation of a line of best fit. Most used the averages in calculations, a more accurate method than reading points from the line. Many well-explained, fully correct, answers were seen.

It was in **parts (d)** and **(e)** that there were limitations. Some candidates seemed unsure of the order in which the equation and the pie chart were to be used, and a lot of work was seen with decimal numbers of fish.

Cambridge Ordinary Level
4040 Statistics November 2024
Principal Examiner Report for Teachers

This was another instance in the paper where a little reflection on the impossibility of some presented answers should have caused the candidate to think again, to review their work, and to try to find and correct the error. It is certainly not expected that candidates will have detailed knowledge of the number of fish ospreys eat, but there should have been some appreciation, from the trend in the collected data, of what might be a reasonable value for week 16. If the number of fish has increased from 4 to 21 over a 10 week period, it is totally unreasonable to expect it to have reached several hundred, or even several thousand, after only another 6 weeks.

Many answers to **part (e)** were speculative about possible things that could happen to the ospreys and the fish, instead of referring to what in Statistics are the usual uncertainties associated with extrapolating results into the future.

STATISTICS

Paper 4040/13
Paper 1

Key messages

Understanding the purpose of calculating a statistical measure is as important as actually being able to calculate it.

After obtaining the answer to a question the candidate should consider whether or not the result is reasonable for the practical situation of the question.

In answering the writing part of a question the candidate should understand properly the meaning of anything they may have learned before using it, and where necessary adapt it to the specific context and requirements of the question.

General comments

The quality of work involving routine numerical calculations was again generally high. This was particularly true of finding crude and standardised rates (see **Question 8** below), calculating the mean and standard deviation of a grouped frequency distribution (see **Question 9** below), and finding the equation of a line of best fit to experimental data (see **Question 11** below). Performance on the short-answer questions was also very good (see **Questions 1, 2 and 3** below).

Amongst the shorter questions it was rare to see full mark answers to the probability question (see **Question 6** below).

Candidates struggled with the writing and 'explain' questions. Often in their attempts at these questions responses were so vague as to have little meaning. The impression was gained that short phrases had been memorised, without full understanding of their meaning, to be produced in the hope of being what was required (see **Questions 8(d), 9(a) and 10(d)** below).

It needs to be emphasised again, as it has been regularly in these reports, that Statistics is a subject with context, used to analyse real-life situations; it is not just about manipulating abstract numbers. Candidates should think whether or not an answer is reasonable for a given situation. There were instances in this paper where unrealistic answers were presented, answers which should have given the candidate at least pause for thought (see **Questions 10(c) and 11(d)** below).

Comments on specific questions

Question 1

The pictogram was interpreted accurately by almost all candidates. The occasional error occurred in part **(b)**, where the answer given was for the number of women who were satisfied, rather than the difference between the number who were dissatisfied and this value.

Question 2

This also was answered well, with only an occasional error occurring in **part (a)(ii)**, resulting from the data not having been arranged in numerical order before identifying the two middle values.

Question 3

Almost all candidates were able to interpret the two-way table correctly. Where an error occurred it was usually in **part (e)**, with the denominator of the probability fraction being given as 40 rather than 14.

Question 4

Good answers to **part (a)** made a simple statement like 'as one variable increases the other variable also increases'. Nothing more than this was required. Answers which attempted to describe the pattern of plots on a scatter diagram showing positive correlation were often unclear and rarely satisfactory. It was also not enough to state only that the variables are directly proportional to each other. Whilst it is true that directly proportionality implies positive correlation, the converse is not the case.

Many completely correct answers to **part (b)** were seen. Where they were not it was usually because of the words 'weak' and 'strong' being omitted.

Question 5

Answers to this question showed most limitations in responses to **parts (a)(iii)** and **(a)(iv)**. Presumably because 'green', 'small' and 'special' were not seen on the diagram, many answers of zero were seen here. Such candidates had not been able, for example, to equate 'green' with not black. Candidates who were able to make such a deduction, there being only two categories for each of colour, size and type, almost always had fully correct answers.

Question 6

Correct deductions in **part (a)** were seen fairly often. Working showing the three possibilities for one cheese and two tomato, with products of three without replacement probabilities, produced the given answer. Incorrect contrived calculations resulting in the value $10/21$ were not only not credited, but acted to the disadvantage of the candidate: it was consequently not possible to extend the correct work which should have been done here to the other possible selections in **part (b)**.

Fully correct answers to **part (b)** were quite rare, though two or three marks were sometimes obtained by candidates who had been successful in **part (a)**. The main reason for this was that the possibility of selecting no cheese sandwiches was frequently overlooked.

Question 7

There was an overall improvement this time in the interpretation of histograms. Most candidates recognised that group frequencies could not necessarily be found by simply reading off the heights of the different bars: the labelling of the vertical axis was crucial in finding these frequencies. So even when answers to **parts (a)** and **(b)** were incorrect, some credit was given provided the values presented were in the right ratio (for example, if all the values presented were twice the correct values).

There were very few satisfactory answers to **part (c)**. Whilst many candidates could point out that a single bar only was used in each of the parts of **part (a)**, and that two bars were used in **part (b)**, this was not considered enough. The essential point was that frequencies for the complete single bars in **part (a)** were known, whilst the frequencies for the *parts* of the two bars in **part (b)** were unknown.

Question 8

Candidates continue to perform well in the routine calculations associated with crude and standardised rates, and full marks were frequently obtained in the first three parts of this fertility rates question. Given the facility with such calculations however, it was regrettable that not all were able to explain in **part (d)** the actual purpose of finding a standardised rate. A few candidates, not giving proper attention to the context, and presumably memorising work on past problems, referred to the health of towns and deaths.

In **part (e)** candidates had to select the appropriate rates to perform the necessary calculations, and performed exceptionally well in almost universally choosing the crude over the standardised. It has however to be pointed out that the question asked for the number of births; a good number of answers entered only the town in the answer line.

Question 9

The main problem with many answers to **part (a)** was that they tended to be so vague and general as to have little meaning. Responses which referred only to making things easier, quicker or being 'time-consuming' were not accepted. One of the best responses for a disadvantage, concise and clear, which many candidates produced, was 'the original data is lost'.

In contrast, answers to **parts (b)** and **(c)** were very good. In **part (c)** working was set out clearly leading to correct answers. When computing the standard deviation from this grouped frequency distribution many more candidates than has sometimes been the case in the past used the formula with Σfx^2 rather than that with $\Sigma f(x - \bar{x})^2$. From this they derived the benefit of not having to work with numbers having four decimal places.

Answers to **parts (d)** and **(e)** were more varied in quality. Candidates who realised in **part (d)** that the two different charges comprising the electricity bill had to be multiplied by different values, one by Σfx and the other by Σf , usually earned full marks. Most errors occurred in finding the charge due to electricity consumption: the \$0.18 was sometimes multiplied by 50, or the sum of the class mid-points, or even the class widths. An incorrect first step that was seen quite often was to add the \$0.18 and \$0.25 before carrying out any multiplication. A correct answer to **part (d)** was almost always followed by a correct answer to **part (e)**.

Question 10

At the outset it needs to be pointed out, as it has been for similar questions in the past, that care has to be taken in the use of the correct total frequency. Although it did not happen often, there were candidates who worked with the maximum value on the cumulative frequency axis (80), and not the total number of deliveries (72). This was in spite of the fact that the latter value could be seen on the graph and was also stated in the question.

Throughout the question those candidates who demonstrated their thinking by drawing lines on the graph are to be commended. Many did this and it enabled examiners to consider awarding marks for method where the numerical value of an answer might be incorrect.

There was a good number of accurate answers to **parts (a)** and **(b)**. The least well answered in these early questions was **part (a)(iv)**: in some answers the value 123 was incorporated into the calculation; in others the cumulative frequency corresponding to 123 km/h was found, the essential first step, but nothing more was done.

It was in **part (c)** where even those candidates with full marks to this point often fell into error. Instead of reading the cumulative frequency corresponding to 130 km/h it was read for 120 km/h. This resulted in an incorrect answer of about 55% being given. At this point, a little reflection on the nature of the situation should have alerted the candidate to the fact that this value was totally unrealistic. In **part (b)(i)** it was established that 29% of Buraid's deliveries were classified as 'fast', so now, after a reduction in their speed, that percentage had to be smaller, not larger.

In **part (d)**, most candidates were able to say that the interquartile range would be unchanged, but explanations as to why this was so were often inadequate. Short phrases of a very general nature, not related to the context, were often given, such as 'it is not affected by addition and subtraction'. Apart from the fact that there was no 'addition' (= increase in delivery speed here), such an answer made no reference, for example, to how the quartiles would change with deliveries 5 km/h slower, or how the amount of dispersion in the distribution would change with all data values 5 km/h smaller.

Question 11

In some cases the graphical work in **parts (a)** and **(b)** could have been improved by making the data point plotting clearer, and the averages plotted in a different way from the data points (for example, using crosses then circled points). But overall, very good work was seen in the first three parts of this question, with candidates showing sound knowledge on how to find the equation of a line of best fit. Most used the averages in calculations, a more accurate method than reading points from the line. Many well-explained, fully correct, answers were seen.

It was in **parts (d)** and **(e)** that there were limitations. Some candidates seemed unsure of the order in which the equation and the pie chart were to be used, and a lot of work was seen with decimal numbers of fish.

This was another instance in the paper where a little reflection on the impossibility of some presented answers should have caused the candidate to think again, to review their work, and to try to find and correct the error. It is certainly not expected that candidates will have detailed knowledge of the number of fish ospreys eat, but there should have been some appreciation, from the trend in the collected data, of what might be a reasonable value for week 16. If the number of fish has increased from 4 to 21 over a 10 week period, it is totally unreasonable to expect it to have reached several hundred, or even several thousand, after only another 6 weeks.

Many answers to **part (e)** were speculative about possible things that could happen to the ospreys and the fish, instead of referring to what in Statistics are the usual uncertainties associated with extrapolating results into the future.

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 questions that begin 'Show ...' and 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. Scales on diagrams should be read carefully, and they should always be labelled, including with any units. The context of a question should be carefully considered, and interpretation of diagrams or calculations should always be made with this context in mind.

General comments

There were two questions on this paper asking candidates to show whether or not a particular statement was true, **Questions 2(b)(ii)** and **8(e)**. It was pleasing to see some well-structured responses, but also there were cases where essential working was missing. In multistage problems, such as **Questions 4(a)** and **5(b)**, clearly set out work enabled examiners to award partial credit if errors had been made.

In **Question 9(f)** final answers of approximately 2.375, were quite common for the estimate of the company's profit in quarter 3 of 2024. Consideration as to whether this answer seemed sensible in the context of the question might have led more candidates to remember to multiply by \$100 000.

In **Question 7(b)**, plots were usually accurate, and axes appropriately labelled with units, although occasionally the vertical axis was labelled cumulative frequency rather than frequency.

Careful consideration of the context is always important. In **Question 1(a)** it was significant in deciding if a measure was a variable or not. In **Question 6a** the context of a race was crucial in judging whether a higher or lower scaled time represented a better result.

Comments on specific questions

Question 1

- (a) Most candidates correctly identified the favourite sport of each person as a qualitative variable and the distance each person travelled as a continuous quantitative variable. The most common error was for the number of people who attended the session to be considered a variable. In the context of the question, where data was being collected about one fitness session, the number attending that session is not a variable.
- (b) Candidates found this the most difficult part of this question. Some, who correctly gave the lower class boundary as 7, often gave the upper class boundary incorrectly as 8. It might have been helpful for these candidates to consider the next class and think about the fact that the upper class boundary of one class is equal to the lower class boundary of the next. Others gave boundaries that would have been correct if rounding to the nearest lap, rather than counting the number of completed laps.
- (c) Candidates were much more successful with finding the class boundaries where rounding to the nearest minute had taken place, with many fully correct answers seen.

Question 2

- (a) Many fully correct solutions were seen in **part (a)**. The most common error occurred when candidates confused the concepts of mutually exclusive and independent events, treating events A and B as if they were independent throughout. A small number got just one probability correct or confused the two probabilities, reversing the answers.
- (b)(i) As with **part (a)**, it was common for candidates to assume that events A and C were independent and multiply the probabilities. A small number used the addition rule correctly but made an arithmetic error or used $P(B)$ rather than $P(C)$ in their calculation.
- (ii) Only the most able candidates were able to provide all the working required to show that events A and C were not independent. It was pleasing to see good communication in those scripts. There were two approaches seen: candidates either assumed independence and compared the resulting $P(A \cap C)$ with their answer to **part (b)(i)**, or they assumed independence and compared the resulting $P(A \cup C)$ with the given value of 0.52. Some, who chose the latter approach, were able to prove that A and C were not independent and score full marks in this part, even if they had assumed independence in **part (b)(i)**. Some candidates were able to score some of the marks by beginning the solution with a correct statement or statements.

Question 3

- (a) Most candidates were able to identify that the first of the diagrams could be used to answer this part. A small number made an error with reading the scale, but many fully correct solutions were seen. A small number tried to use the percentage sectional bar chart, but this approach would only work if combined with information from the other two diagrams, and this was not seen.
- (b) This part was more difficult as it required information to be obtained from both the box-and-whisker diagram and the sectional bar chart. Some candidates correctly identified from the box-and-whisker diagram that 25 per cent of the women were under 30, but they sometimes found 25 per cent of 140, rather than of the 80 women. Many candidates simply subtracted the lowest age for women of 16 from the lower quartile value of 30, giving 14 as the answer.
- (c) This final part was the most difficult, requiring information to be obtained from all three diagrams. Only the most able candidates were able to successfully combine the information presented. Many did however use the percentage sectional bar chart and used 50 per cent in their final calculation, but they often found 50 per cent of 140, rather than using the other diagrams to find out how many people of each gender were aged 50 and over.

Question 4

- (a) It was pleasing to see many fully correct solutions to **part (a)**. An algebraic approach was used in most cases. A small number did not know how to begin, but many multiplied weights by price relatives, even if they were unable to proceed further.
- (b) It was very common to award one mark out of two in this question. The most common error seen was for the overall cost to be incorrectly described as the expenditure. Most candidates correctly found the amount of the percentage decrease and correctly indicated the years between which this change had taken place.
- (c) Many candidates made incorrect references to changes in prices rather than changes in amounts. These changes in amounts needed to be given in the context of the photographer. The most commonly seen correct suggestions were changes to the amount of equipment and changes to distances travelled.

Question 5

- (a) Most candidates made a good attempt at this question. It was largely unstructured, and so the first step was to work out the possible prizes. The blank table provided a clue as to the number of possible prizes, and many candidates found these successfully. Some candidates drew themselves pictures of the two bags to help with visualising the situation. Some candidates got the probabilities of prizes of \$2 and \$4 correct, but made an error with the prize of \$3, not correctly thinking about

the two ways in which that prize could be obtained. In these cases, their probability distributions contained probabilities that did not sum to 1. This fact should have alerted them to the existence of an error.

- (b) Most candidates scored at least one or two of the available marks by correctly summing the products of their prizes and corresponding probabilities. Multiplication by 15 was often the next step. Some candidates omitted to consider the charge of \$2.50 and some did not retain sufficient accuracy during the calculation. It was necessary, for the final mark, to state whether the amount obtained represented a profit or a loss for Yuri, and this was sometimes missing.

Question 6

- (a) Correct scaled scores of -2 and -1.25 for the 100 m and 200 m races, respectively, were seen on many scripts. Based on this, many candidates suggested that Amna should enter the 200 m race rather than concluding that, in the context of times to complete a race, a lower scaled score was better, and that therefore she should enter the 100 m race.
- (b) The most common error in **part (b)** was for the standard deviation to be treated in the same way as the mean and thus an answer for the standard deviation of 5.6 was commonly seen. There were, however, many fully correct answers to this part.

Question 7

- (a) Most candidates scored full marks in **part (a)**, with the most common error being a final entry of 0 in the cumulative frequency column.
- (b) Frequency polygons were usually accurately drawn, and it was pleasing to see that axes were usually labelled, and units included in those labels. The most common error was for the plots to be made at the upper boundary of the interval, rather than at the midpoint, suggesting some confusion with cumulative frequency polygons. Straight line segments were usually used correctly to connect the plots.
- (c) Candidates often successfully found an estimate for the median using linear interpolation. A few candidates confused the figures for the frequencies and the cumulative frequencies in their calculations. A small number of candidates incorrectly added 1 to 76 before dividing by 2 to find the position of the median, treating the data as if it were ungrouped discrete data.
- (d) The final part of this question was more difficult, but many fully correct solutions were seen, with most candidates beginning by doubling the given width. An alternative, correct, approach of halving the lengths in the table was also sometimes seen. Some candidates treated the given width as if it were a length, and confusion between the frequencies and cumulative frequencies was more prevalent in this part.

Question 8

- (a) A range of correct responses were seen suggesting a good understanding of the disadvantages of open questions, with many pointing to the difficulty of analysis. Incorrect responses seen often referred to bias or that the company might not like an opinion that was expressed.
- (b) Many correct answers were seen, but incorrect answers were also common. Incorrect responses included 'questionnaire' and the names of different types of sampling method.
- (c) This question was well answered by most candidates. The most common correct advantage of a sample was that it would be quicker to conduct. It was not sufficient to say simply that it was easier, without providing an explanation, such as easier to collect the data, or easier to analyse the data. There were also many correct disadvantages seen, with the fact that it may not be representative being the most common.
- (d) Many candidates found a correct systematic sample starting at 02. Common errors were for a correct interval size of 11, but an incorrect starting value, such as 27, with such samples often resulting in some values beyond 65, so out of the range of the data. Some found systematic samples, but with incorrect interval sizes, such as 10 or 6, and others found simple random samples.

- (e) The most successful candidates were able to communicate their reasoning clearly and show that the sample was not representative in terms of contract type, but some solutions were missing important aspects. Some candidates correctly identified their sample members as either full or part-time and others provided calculations to show how many full or part-time members a representative sample would contain, but often one or other of these essential stages was missing. A common error was for candidates to check whether the sample was representative in terms of age group and contract type rather than age group alone.
- (f) Most candidates answered this part well. A few candidates reached an answer of 4 from incorrect working, most often by calculating $\frac{23}{66} \times 11 = 3.833\dots$ and then rounding to 4.
- (g) Only the most able candidates gave the three possible sample sizes. Some candidates worked out one correct sample size, usually 12, but did not consider the situations when rounding to the nearest integer would take place. Some candidates appeared to give answers that were strata sizes rather than sizes of the total sample.

Question 9

- (a) Many candidates correctly identified that even-pointed moving averages were the ones that needed centring. It was much rarer to see the explanation as to why centring was needed in those cases, namely so that the moving average values would coincide with original readings. Many candidates instead went on to explain the purpose of calculating seasonal components rather than how centring made it possible to calculate them.
- (b) Most candidates were able to centre the given moving average values. Some candidates divided the appropriate sum by 4 rather than 2, and others calculated 4-point averages of the given 4-point moving averages, rather than centring them.
- (c) A clear understanding of what was required to calculate a seasonal component was seen from many candidates. At least one appropriate subtraction was often seen. Some candidates, having added two appropriate differences, omitted to divide by two. Some subtracted the wrong way around, perhaps for ease of calculation, but did not adjust the sign at the end.
- (d) Only the most able candidates correctly interpreted the seasonal component. The most common error was to suggest that it showed that profits in quarter 3 were decreasing, rather than that it showed how values in quarter 3 compared to the moving average or trend line values.
- (e) Plots tended to be accurate, with just a small number of candidates misreading the scale. The trend line was usually drawn in an appropriate place, although sometimes it was drawn too steeply, with the first three plots well below the line and the last three plots well above the line. Some lines were drawn too short, so that they could not be used in **part (f)**.
- (f) Where trend lines had been extended as far as quarter 3 of 2024, readings from these trend lines were usually accurate. As in previous years, some candidates simply gave that reading as their final answer and achieved no marks. When the seasonal component was included in the calculation it was usually added, although a small number of candidates incorrectly subtracted their value. To achieve full marks in this final part of the question, it was necessary to give a final answer that took account of the fact that the profits were in \$100 000s. This aspect was often missing.

Question 10

- (a) Most candidates found the correct probability in **part (a)**, with just a very small number totalling up the number of birds for the denominator, rather than totalling the number of nests.
- (b)(i) Correct calculations were also usually seen in **part (b)(i)**, with just a small number of candidates making an arithmetic slip, such as $0 \times 4 = 4$.
- (ii) Fully correct solutions were often seen, with the most common incorrect response being $\frac{8}{42} = \frac{4}{21}$.

Cambridge Ordinary Level
4040 Statistics November 2024
Principal Examiner Report for Teachers

- (c) Most candidates were able to gain some marks on this question, but errors in the table were quite common. The most common of those errors was an incorrect range, often given as 21.
- (d) Some candidates gained one of the two available marks, but it was rare to see responses that earned both marks. One mark was often awarded for a comment about the range appearing to support the claim. Other candidates earned a mark by commenting that the interquartile range or the standard deviation were not available, but without reference to the range. It was rare to see either the disadvantage of the range, namely that it can be affected by extreme values, or a comment that the standard deviation, for example, would have been a better measure of spread than the range. Some candidates incorrectly thought that the mean values supported the claim.
- (e) Many candidates correctly found the combined mean. The most common incorrect answer was 1.95, coming from adding the two means together and dividing by 2, without consideration of the different numbers of nests in each section.
- (f) Most candidates were able to gain some marks in this part, but it was rare to award all 3 marks. Often a response of D, there is not enough information to know the effect, was seen in the final situation.

STATISTICS

Paper 4040/23
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 questions that begin 'Show ...' and 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. Scales on diagrams should be read carefully, and they should always be labelled, including with any units. The context of a question should be carefully considered, and interpretation of diagrams or calculations should always be made with this context in mind.

General comments

There were two questions on this paper asking candidates to show whether or not a particular statement was true, **Questions 2(b)(ii)** and **8(e)**. It was pleasing to see some well-structured responses, but also there were cases where essential working was missing. In multistage problems, such as **Questions 4(a)** and **5(b)**, clearly set out work enabled examiners to award partial credit if errors had been made.

In **Question 9(f)** final answers of approximately 2.375, were quite common for the estimate of the company's profit in quarter 3 of 2024. Consideration as to whether this answer seemed sensible in the context of the question might have led more candidates to remember to multiply by \$100 000.

In **Question 7(b)**, plots were usually accurate, and axes appropriately labelled with units, although occasionally the vertical axis was labelled cumulative frequency rather than frequency.

Careful consideration of the context is always important. In **Question 1(a)** it was significant in deciding if a measure was a variable or not. In **Question 6a** the context of a race was crucial in judging whether a higher or lower scaled time represented a better result.

Comments on specific questions

Question 1

- (a) Most candidates correctly identified the favourite sport of each person as a qualitative variable and the distance each person travelled as a continuous quantitative variable. The most common error was for the number of people who attended the session to be considered a variable. In the context of the question, where data was being collected about one fitness session, the number attending that session is not a variable.
- (b) Candidates found this the most difficult part of this question. Some, who correctly gave the lower class boundary as 7, often gave the upper class boundary incorrectly as 8. It might have been helpful for these candidates to consider the next class and think about the fact that the upper class boundary of one class is equal to the lower class boundary of the next. Others gave boundaries that would have been correct if rounding to the nearest lap, rather than counting the number of completed laps.
- (c) Candidates were much more successful with finding the class boundaries where rounding to the nearest minute had taken place, with many fully correct answers seen.

Question 2

- (a) Many fully correct solutions were seen in **part (a)**. The most common error occurred when candidates confused the concepts of mutually exclusive and independent events, treating events A and B as if they were independent throughout. A small number got just one probability correct or confused the two probabilities, reversing the answers.
- (b)(i) As with **part (a)**, it was common for candidates to assume that events A and C were independent and multiply the probabilities. A small number used the addition rule correctly but made an arithmetic error or used $P(B)$ rather than $P(C)$ in their calculation.
- (ii) Only the most able candidates were able to provide all the working required to show that events A and C were not independent. It was pleasing to see good communication in those scripts. There were two approaches seen: candidates either assumed independence and compared the resulting $P(A \cap C)$ with their answer to **part (b)(i)**, or they assumed independence and compared the resulting $P(A \cup C)$ with the given value of 0.52. Some, who chose the latter approach, were able to prove that A and C were not independent and score full marks in this part, even if they had assumed independence in **part (b)(i)**. Some candidates were able to score some of the marks by beginning the solution with a correct statement or statements.

Question 3

- (a) Most candidates were able to identify that the first of the diagrams could be used to answer this part. A small number made an error with reading the scale, but many fully correct solutions were seen. A small number tried to use the percentage sectional bar chart, but this approach would only work if combined with information from the other two diagrams, and this was not seen.
- (b) This part was more difficult as it required information to be obtained from both the box-and-whisker diagram and the sectional bar chart. Some candidates correctly identified from the box-and-whisker diagram that 25 per cent of the women were under 30, but they sometimes found 25 per cent of 140, rather than of the 80 women. Many candidates simply subtracted the lowest age for women of 16 from the lower quartile value of 30, giving 14 as the answer.
- (c) This final part was the most difficult, requiring information to be obtained from all three diagrams. Only the most able candidates were able to successfully combine the information presented. Many did however use the percentage sectional bar chart and used 50 per cent in their final calculation, but they often found 50 per cent of 140, rather than using the other diagrams to find out how many people of each gender were aged 50 and over.

Question 4

- (a) It was pleasing to see many fully correct solutions to **part (a)**. An algebraic approach was used in most cases. A small number did not know how to begin, but many multiplied weights by price relatives, even if they were unable to proceed further.
- (b) It was very common to award one mark out of two in this question. The most common error seen was for the overall cost to be incorrectly described as the expenditure. Most candidates correctly found the amount of the percentage decrease and correctly indicated the years between which this change had taken place.
- (c) Many candidates made incorrect references to changes in prices rather than changes in amounts. These changes in amounts needed to be given in the context of the photographer. The most commonly seen correct suggestions were changes to the amount of equipment and changes to distances travelled.

Question 5

- (a) Most candidates made a good attempt at this question. It was largely unstructured, and so the first step was to work out the possible prizes. The blank table provided a clue as to the number of possible prizes, and many candidates found these successfully. Some candidates drew themselves pictures of the two bags to help with visualising the situation. Some candidates got the probabilities of prizes of \$2 and \$4 correct, but made an error with the prize of \$3, not correctly thinking about

the two ways in which that prize could be obtained. In these cases, their probability distributions contained probabilities that did not sum to 1. This fact should have alerted them to the existence of an error.

- (b) Most candidates scored at least one or two of the available marks by correctly summing the products of their prizes and corresponding probabilities. Multiplication by 15 was often the next step. Some candidates omitted to consider the charge of \$2.50 and some did not retain sufficient accuracy during the calculation. It was necessary, for the final mark, to state whether the amount obtained represented a profit or a loss for Yuri, and this was sometimes missing.

Question 6

- (a) Correct scaled scores of -2 and -1.25 for the 100 m and 200 m races, respectively, were seen on many scripts. Based on this, many candidates suggested that Amna should enter the 200 m race rather than concluding that, in the context of times to complete a race, a lower scaled score was better, and that therefore she should enter the 100 m race.
- (b) The most common error in **part (b)** was for the standard deviation to be treated in the same way as the mean and thus an answer for the standard deviation of 5.6 was commonly seen. There were, however, many fully correct answers to this part.

Question 7

- (a) Most candidates scored full marks in **part (a)**, with the most common error being a final entry of 0 in the cumulative frequency column.
- (b) Frequency polygons were usually accurately drawn, and it was pleasing to see that axes were usually labelled, and units included in those labels. The most common error was for the plots to be made at the upper boundary of the interval, rather than at the midpoint, suggesting some confusion with cumulative frequency polygons. Straight line segments were usually used correctly to connect the plots.
- (c) Candidates often successfully found an estimate for the median using linear interpolation. A few candidates confused the figures for the frequencies and the cumulative frequencies in their calculations. A small number of candidates incorrectly added 1 to 76 before dividing by 2 to find the position of the median, treating the data as if it were ungrouped discrete data.
- (d) The final part of this question was more difficult, but many fully correct solutions were seen, with most candidates beginning by doubling the given width. An alternative, correct, approach of halving the lengths in the table was also sometimes seen. Some candidates treated the given width as if it were a length, and confusion between the frequencies and cumulative frequencies was more prevalent in this part.

Question 8

- (a) A range of correct responses were seen suggesting a good understanding of the disadvantages of open questions, with many pointing to the difficulty of analysis. Incorrect responses seen often referred to bias or that the company might not like an opinion that was expressed.
- (b) Many correct answers were seen, but incorrect answers were also common. Incorrect responses included 'questionnaire' and the names of different types of sampling method.
- (c) This question was well answered by most candidates. The most common correct advantage of a sample was that it would be quicker to conduct. It was not sufficient to say simply that it was easier, without providing an explanation, such as easier to collect the data, or easier to analyse the data. There were also many correct disadvantages seen, with the fact that it may not be representative being the most common.
- (d) Many candidates found a correct systematic sample starting at 02. Common errors were for a correct interval size of 11, but an incorrect starting value, such as 27, with such samples often resulting in some values beyond 65, so out of the range of the data. Some found systematic samples, but with incorrect interval sizes, such as 10 or 6, and others found simple random samples.

- (e) The most successful candidates were able to communicate their reasoning clearly and show that the sample was not representative in terms of contract type, but some solutions were missing important aspects. Some candidates correctly identified their sample members as either full or part-time and others provided calculations to show how many full or part-time members a representative sample would contain, but often one or other of these essential stages was missing. A common error was for candidates to check whether the sample was representative in terms of age group and contract type rather than age group alone.
- (f) Most candidates answered this part well. A few candidates reached an answer of 4 from incorrect working, most often by calculating $\frac{23}{66} \times 11 = 3.833\dots$ and then rounding to 4.
- (g) Only the most able candidates gave the three possible sample sizes. Some candidates worked out one correct sample size, usually 12, but did not consider the situations when rounding to the nearest integer would take place. Some candidates appeared to give answers that were strata sizes rather than sizes of the total sample.

Question 9

- (a) Many candidates correctly identified that even-pointed moving averages were the ones that needed centring. It was much rarer to see the explanation as to why centring was needed in those cases, namely so that the moving average values would coincide with original readings. Many candidates instead went on to explain the purpose of calculating seasonal components rather than how centring made it possible to calculate them.
- (b) Most candidates were able to centre the given moving average values. Some candidates divided the appropriate sum by 4 rather than 2, and others calculated 4-point averages of the given 4-point moving averages, rather than centring them.
- (c) A clear understanding of what was required to calculate a seasonal component was seen from many candidates. At least one appropriate subtraction was often seen. Some candidates, having added two appropriate differences, omitted to divide by two. Some subtracted the wrong way around, perhaps for ease of calculation, but did not adjust the sign at the end.
- (d) Only the most able candidates correctly interpreted the seasonal component. The most common error was to suggest that it showed that profits in quarter 3 were decreasing, rather than that it showed how values in quarter 3 compared to the moving average or trend line values.
- (e) Plots tended to be accurate, with just a small number of candidates misreading the scale. The trend line was usually drawn in an appropriate place, although sometimes it was drawn too steeply, with the first three plots well below the line and the last three plots well above the line. Some lines were drawn too short, so that they could not be used in **part (f)**.
- (f) Where trend lines had been extended as far as quarter 3 of 2024, readings from these trend lines were usually accurate. As in previous years, some candidates simply gave that reading as their final answer and achieved no marks. When the seasonal component was included in the calculation it was usually added, although a small number of candidates incorrectly subtracted their value. To achieve full marks in this final part of the question, it was necessary to give a final answer that took account of the fact that the profits were in \$100 000s. This aspect was often missing.

Question 10

- (a) Most candidates found the correct probability in **part (a)**, with just a very small number totalling up the number of birds for the denominator, rather than totalling the number of nests.
- (b)(i) Correct calculations were also usually seen in **part (b)(i)**, with just a small number of candidates making an arithmetic slip, such as $0 \times 4 = 4$.
- (ii) Fully correct solutions were often seen, with the most common incorrect response being $\frac{8}{42} = \frac{4}{21}$.

Cambridge Ordinary Level
4040 Statistics November 2024
Principal Examiner Report for Teachers

- (c) Most candidates were able to gain some marks on this question, but errors in the table were quite common. The most common of those errors was an incorrect range, often given as 21.
- (d) Some candidates gained one of the two available marks, but it was rare to see responses that earned both marks. One mark was often awarded for a comment about the range appearing to support the claim. Other candidates earned a mark by commenting that the interquartile range or the standard deviation were not available, but without reference to the range. It was rare to see either the disadvantage of the range, namely that it can be affected by extreme values, or a comment that the standard deviation, for example, would have been a better measure of spread than the range. Some candidates incorrectly thought that the mean values supported the claim.
- (e) Many candidates correctly found the combined mean. The most common incorrect answer was 1.95, coming from adding the two means together and dividing by 2, without consideration of the different numbers of nests in each section.
- (f) Most candidates were able to gain some marks in this part, but it was rare to award all 3 marks. Often a response of D, there is not enough information to know the effect, was seen in the final situation.