Qualification Accredited

OCR
Oxford Cambridge and RSA

A LEVEL

Examiners' report

MATHEMATICS A

H240

For first teaching in 2017

H240/02 Summer 2023 series

Version 1

ocr.org.uk/maths

Contents

Introduction

Our examiners' reports are produced to offer constructive feedback on candidates' performance in the examinations. They provide useful guidance for future candidates.

The reports will include a general commentary on candidates' performance, identify technical aspects examined in the questions and highlight good performance and where performance could be improved. A selection of candidate answers is also provided. The reports will also explain aspects which caused difficulty and why the difficulties arose, whether through a lack of knowledge, poor examination technique, or any other identifiable and explainable reason.

Where overall performance on a question/question part was considered good, with no particular areas to highlight, these questions have not been included in the report.

A full copy of the question paper and the mark scheme can be downloaded from OCR.

Paper 2 series overview

The overall standard of candidates' work was broadly similar to 2019, with a slightly lower mean mark overall. There were notably fewer very high marks, meaning that candidate mark frequencies were more closely clustered, with a nearer-normal distribution than in 2019

Most candidates attempted the majority of questions, and many – but not all – candidates demonstrated clear confidence with the assumed knowledge from GCSE and initial AS content. However, a number of candidates did appear to have some gaps in this foundation knowledge (for example, the box and whisker diagram in Question 8 (a)).

As in 2019, some candidates seemed unsure as to how much working they needed to show in any given question (and the guidance on p6 of the 2019 report remains relevant and worth reviewing). In particular, candidates did not appear to pay sufficiently close attention to the instructions 'Determine' and 'Detailed reasoning' which meant that in some cases full marks could not be given, despite the correct answer being obtained. This was particularly apparent in Questions 4 (a) and 5 on this paper.

Many candidates also lost marks because the explanations they provided in written questions (e.g. Question 9 (d) and most of Question 13) were not sufficiently specific to the context of the question. As on H230/01, and this paper in 2019, candidates should be reminded to make sure that their explanations are detailed, in context, and answering the specific request in the question. Some candidates stated a generic fact, or only gave a partial explanation (e.g. without a justification in context) which could not be credited.

Candidates who did well on this paper generally:	Candidates who did less well on this paper generally:
• demonstrated good understanding of both pure (Section A) and statistics (Section B) • applied strong 'pure skills' in algebraic questions, and provided appropriately rigorous proofs when requested • were able to apply their understanding of statistical concepts in problem-solving questions • provided detailed explanations in context • were able to use their knowledge of the Large Data Set to inform their responses to Question 13 • showed appropriate working when required to justify their answers, and made good use of a calculator when necessary.	• showed insufficient working in questions where 'Determine' or 'Detailed reasoning' was used • did not appear to have secure understanding of the assumed knowledge from GCSE • did not show sufficient rigour in 'proof' questions • did not provide sufficiently detailed explanations • misunderstood parts of the Large Data Set question (Question 13) • were only able to answer the simpler probability questions.

OCR support

OCR publish a classroom poster and associated guidance on command words, which centres should make sure candidates are familiar with. These can be found on Teach Cambridge.

Section A overview

Most candidates were able to answer the more straightforward questions in Section A well, though many did not provide sufficient rigour or demonstration of their method in the more complex questions. Candidates should be reminded that although this is a 'Statistics' paper the level of rigour required in Section A (Pure) is the same across the specification, especially where a question instructs 'Show that', 'Determine' or 'Detailed reasoning'.

Question 1 (a) (i)

1 (a) (i) Express $x^2-8x+11$ in the form $(x-a)^2+b$ where a and b are constants. **[2]**

This question was generally well answered by the majority of candidates. A small number of candidates made an error in completing the square. Some candidates went on to mis-state the values of *a* or *b* (without regard to the given form, e.g. stating *a* = −4), but this subsequent working was ignored if the completed square form was given correctly.

Question 1 (a) (ii)

(ii) Hence write down the minimum value of $x^2-8x+11$. **[1]**

This question was much less well answered than (a) (i), even by candidates who had part (i) correct. Many misunderstood the request for 'the minimum value' and gave the minimum point (4,-5) - which was accepted, provided it was correct from their completed square form. Many candidates gave 4 (the x-value at which the minimum occurs) which on its own was not sufficient.

Question 1 (b)

(b) Determine the value of the constant k for which the equation $x^2-8x+11=k$ has two equal roots. **[2]**

Although some candidates may have been able to deduce that *k* = −5 is the required value, the question stated 'Determine' which meant that some working or reasoning was required. The most common method was to use the discriminant, setting $b^2-4ac=0$ and, where attempted, this was generally well done.

Some candidates made an error in solving and obtained *k* = 5, other candidates confused this with the range of values of *k* for which there were any real roots, and lost the final mark by using > instead of =. Unless candidates went on to finish with a statement of equality *k* = −5 this could not obtain the final accuracy mark.

Question 2 (a)

2 The points O and A have position vectors $\begin{pmatrix}0\\0\\0\end{pmatrix}$ and $\begin{pmatrix}6\\0\\8\end{pmatrix}$ respectively. The point P is such that $\overrightarrow{OP} = k\overrightarrow{OA}$, where k is a non-zero constant.

(a) Find, in terms of k, the length of OP. **[1]**

Many candidates found the vector *OP* correctly but did not go on to give the required expression for the length. A few candidates muddled the 6 and 8. Almost all candidates gave the answer as 10*k* which was permitted, even though strictly the correct expression is 10|*k*| since length cannot be negative.

Question 2 (b)

Point B has position vector $\begin{pmatrix}1\\2\\3\end{pmatrix}$ and angle OPB is a right angle.

(b) Determine the value of k. **[4]**

Many candidates appeared to struggle with this question, and a variety of unfruitful methods were seen. The command word is 'Determine' so some working is required. Many candidates misunderstood the 'triangle' to which they were applying Pythagoras and so obtained a 'not useful' equation. Candidates were permitted to 'imply' the B mark for *BP* or *PB* but (in a determine question) it would of course be preferable if this was explicitly stated first.

Algebraic errors were quite commonly seen in the equation derived from Pythagoras, although many of those who completed this correctly went on disregard the additional solution $k = 0$, which is good practice.

In general those candidates using the scalar product method (from Further Maths) obtained the correct answer and full credit. A few candidates recognised that a simpler equation could be obtained by working with the angle *AOB* (which is the same as *BOP*), and either using OA.BP or the right-angled triangle to obtain the required answer.

Question 3

3 In this question you must show detailed reasoning.

Find the exact area of the region enclosed by the curve $y = \frac{1}{x+2}$, the two axes and the line $x = 2.5$. **[3]**

Most candidates answered this question well. A small number did not heed the '**In this question you must show detailed reasoning**' instruction and used their calculator to compute the integral, giving an inexact answer, which could not be credited. Candidates needed to use log rules to combine and simplify their final answer into an exact form.

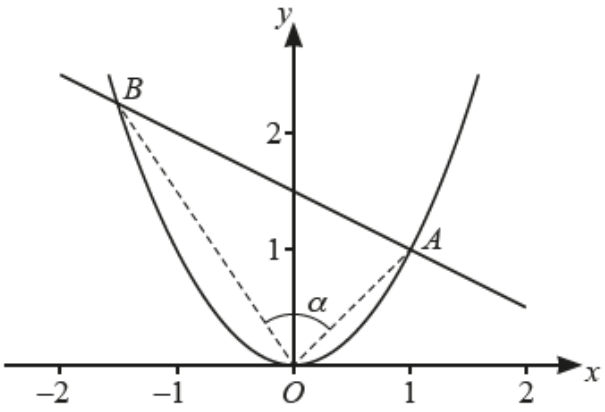
Some candidates did not find the correct limits, which meant that they could only obtain the first M mark (if their integration was correct). A small number used a substitution, generally correctly, although this represented much more work than was required as this integral can be found directly.

Detailed reasoning and exact answers

'**Detailed reasoning**' means that sufficient working must be seen to clearly justify the answer obtained. The use of calculators is permitted to check or verify solutions, but the algebraic working given must 'stand alone' and lead to the required answer. In Question 3, many candidates performed the integration correctly but then went on to use a calculator to give an inexact answer – even though the question specifically requested the 'exact area'.

Question 4 (a)

4 The diagram shows part of the graph of $y = x^2$. The normal to the curve at the point $A(1, 1)$ meets the curve again at B. Angle AOB is denoted by α.

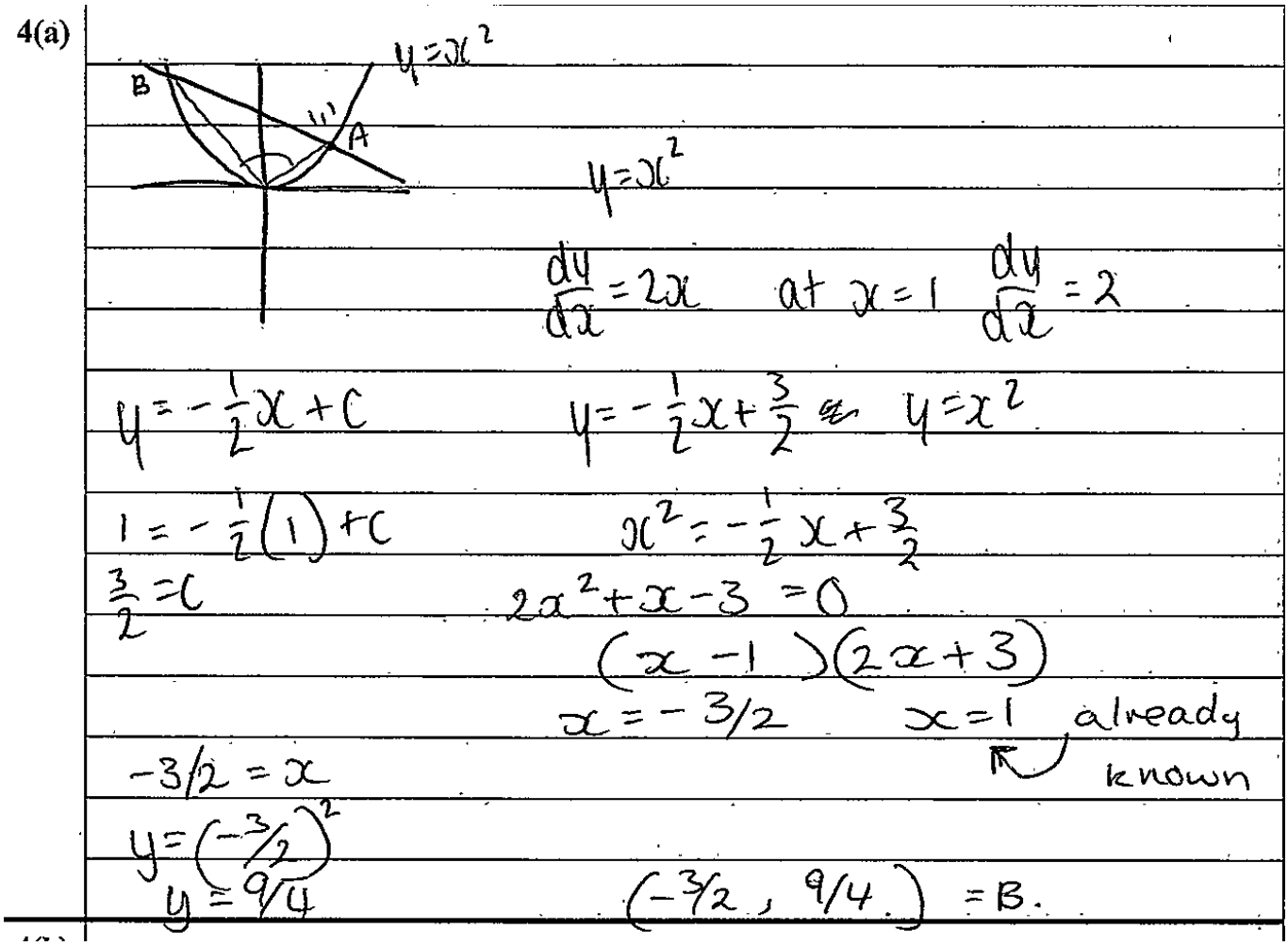


(a) Determine the coordinates of B. **[6]**

This question was very well answered, with the majority of candidates reaching the required answer. However, again some candidates did not heed the 'Determine' instruction and gave minimal working, or an answer only (BC) without working. Where the command word is 'Determine' working is required to support the answer, even if it can be done on a calculator.

Many candidates found the equation of the normal correctly but then did not show any working for how this led to the coordinates of *B*, which meant the final 3 marks could not be given – even for a correct answer.

Exemplar 1



This candidate's response shows all of the required working to justify their (correct) answer. They have shown how the equation of the normal is obtained and then correctly substituted to solve for the coordinates required.

Question 4 (b)

(b) Hence determine the exact value of $\tan\alpha$. **[3]**

This challenging question was not well answered. Many candidates either used their calculator (to compute inverse tan, add the angles together and then 're-tan') which was not appropriate for 'Determine' as they then provided inexact working which did not lead to tan α = 5 exactly. Alternatively, a number of candidates misinterpreted the angle and assumed that BAO=90°, when it isn't (the line is normal to the curve, not the line OA).

The main method in the scheme, using the cosine rule, was most common, although many candidates made a slip in computing one of the lengths. A small number of candidates used the (more direct) method given in the scheme as Alt 1, but on both methods candidates sought shortcuts without showing exact working, which meant they could not be credited.

As in Question 3, the 'exact' instruction in the question is critical here. An inexact answer that rounds to 5 was not sufficient, exact working that leads to an exact answer was required. However, this is a good example of a question where some candidates made good use of their calculator to check their solution (having found it using exact working) by computing α = 78.69…°

Exemplar 2

4(b)

$OA = \sqrt{1+1} = \sqrt{2}$

$OB = \sqrt{\frac{9}{4} + \frac{81}{16}} = \frac{3}{4}\sqrt{13}$

$AB = \sqrt{(-\frac{3}{2}-1)^2 + (\frac{9}{4}-1)^2} = \frac{5}{4}\sqrt{5}$

$\cos\alpha = \frac{OB^2 + OA^2 - AB^2}{2 \cdot OB \cdot OA}$

$= \frac{\frac{117}{16} + 2 - \frac{125}{16}}{2 \times \frac{3}{4}\sqrt{13} \times \sqrt{2}}$

$= \frac{\sqrt{26}}{26}$

$\alpha = 78.69^\circ$

$\tan\alpha = 5.$

This candidate (who obtained full marks) is an example of sufficient exact working being shown, together with the (calculator-obtained) value that has been used to verify their solution. Although included in the flow of the working here, the computation of α does not undermine their correct, exact, working for cos α.

Question 5 (a) (i)

5 In this question you must show detailed reasoning.

The function f is defined by $f(x) = \cos x + \sqrt{3}\sin x$ with domain $0 \leqslant x \leqslant 2\pi$.

(a) Solve the following equations.

(i) $f'(x) = 0$ **[4]**

This question was generally well answered. Some candidates appeared uncertain how to differentiate the given expression, either making errors in differentiation or attempting to (unnecessarily) apply the product rule. Most candidates were able to set their derivative = 0 and obtain an expression from which values could be deduced.

The entirety of Question 5 is **Detailed reasoning** which means that sufficient working is required to support all answers, throughout. There were fewer cases of 'answers without working' in this question - perhaps because most candidates tended to write down the expression and manipulate it directly.

Although the differentiated expression resolves straightforwardly to $\tan x = \sqrt{3}$, some candidates did not spot this and so chose to square in order to combine the terms. This was acceptable (as long as done correctly, some candidates wrote x – y = 0 followed by $x^2 - y^2 = 0$ which is not a correct step of working or squared the entire expression while omitting the 'cross term' which could lead to the correct solutions, from incorrect working), but candidates then obtained additional solutions which they needed to test and eliminate - many did not do so (or eliminated the wrong solutions) and lost the final A mark as a result.

A small number of candidates used $R\cos(x + \alpha)$ to combine the two terms first, which makes the differentiation more straightforward.

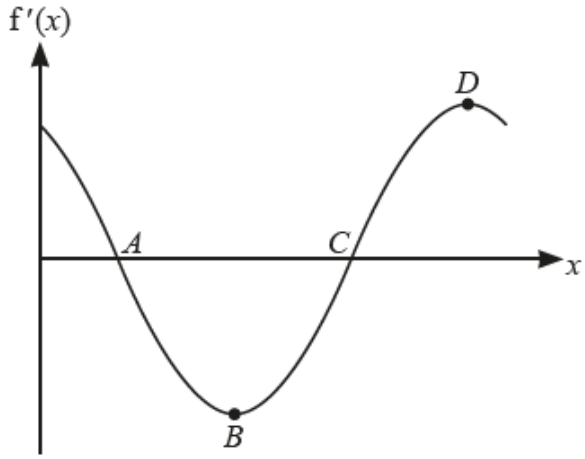
Question 5 (a) (ii)

(ii) $f''(x) = 0$ **[3]**

This was very similarly answered to part (a) (i), with most candidates reaching at least one of the required answers. Again, some squared their equation and then did not eliminate additional solutions within the range.

Question 5 (b)

The diagram shows the graph of the gradient function $y = f'(x)$ for the domain $0 \leqslant x \leqslant 2\pi$.



(b) Use your answers to parts **(a)(i)** and **(a)(ii)** to find the coordinates of points A, B, C and D. **[2]**

Relatively few candidates gave a fully correct set of answers to this question. Many candidates were able to correctly match their x-values from part (a) to the four points, but then didn't go on to compute the corresponding y-values (receiving B1 under the SC if all four matched correctly). A number of candidates also assumed, incorrectly, that the y-values were +/-1.

Misconception

Where a question asks for 'coordinates', candidates should provide coordinate pairs (x,y) – when working in 2 dimensions.

Question 5 (c) (i)

(c) (i) Explain how to use the graph of the gradient function to find the values of x for which $f(x)$ is increasing. **[1]**

This question proved challenging for many candidates. Some were able to state the correct generic fact ($f'(x) > 0$) but without reference to the graph this was not sufficient for the 'Explain' required.

Some candidates misinterpreted the setup and referred to $f''(x)$ or where the gradient of this graph was positive or negative, or concave up/down. Some candidates referred to the points *A* and *C* although this was not required. Candidates should make sure that they provide an explanation that is specific to the request in the question, not just stating a generic fact.

A few candidates also mis-stated what to look for, e.g. 'where the graph is above the *y*-axis', or wrongly assumed that the graph was of $f(x)$.

Assessment for learning

Where a question asks candidate to 'Explain' they should make sure that they have provided the detailed explanation required, in the specific context of the question. Stating a generic fact will not normally be sufficient. In this case, candidates needed to 'explain **how to use the graph'**, so simply stating the definition of an increasing function was not sufficient.

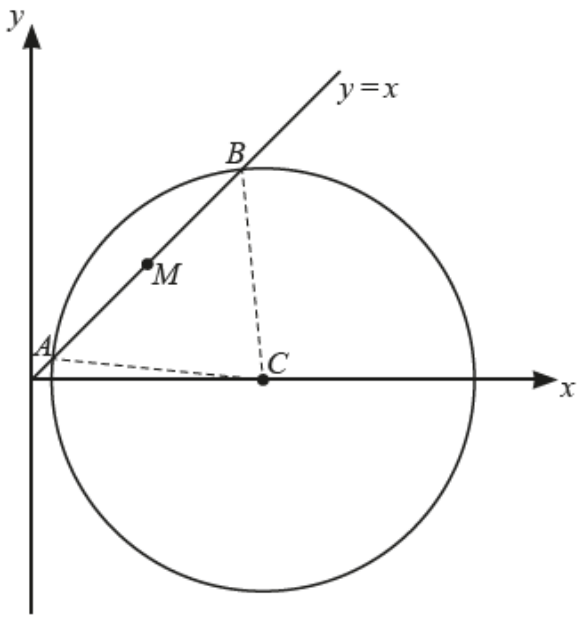
Question 5 (c) (ii)

(ii) Using set notation, write down the set of values of x for which $f(x)$ is increasing in the domain $0 \leqslant x \leqslant 2\pi$. **[2]**

There were 2 marks available here. The first was for identifying the correct range of *x*-values (regardless of set notation used). Very few candidates managed to do this correctly, even having obtained the correct values in part (a) and 'matched' them in part (b) – with many omitting the outer limits of both ranges. The second mark (which was available independently of the first) was for giving their range in correct set notation, represented as the Union of two sets. More candidates obtained this mark than the first, but many made errors such as ∩ instead of ∪, or omitting '*x*:' from one or both sets, or otherwise incorrect set notation.

Question 6 (a)

6 A circle has centre C which lies on the x-axis, as shown in the diagram. The line $y = x$ meets the circle at A and B. The midpoint of AB is M.



The equation of the circle is $x^2 - 6x + y^2 + a = 0$, where a is a constant.

(a) In this question you must show detailed reasoning.

Show that the area of triangle ABC is $\frac{3}{2}\sqrt{9-2a}$. **[7]**

Many candidates appeared to find this question very challenging. A good number of candidates were able to obtain the first 1-2 marks by either attempting to complete the square or substitute in $y = x$ and solve for one or both coordinates of the points of intersection A and B. Many candidates started with the alternative method (attempting to complete the square) and then reverted to the main method.

Those candidates who proceeded beyond that first stage generally went on to successfully reach (or almost reach) the given answer - deducing the coordinates of M and finding two of the required lengths to allow them to form an expression for the area of the triangle. Many of these correctly handled the length of AB and were able to write down a correct expression for the required area.

There was (despite the **Detailed reasoning** instruction) some evidence of candidates attempting to reverse-engineer the given result from 'similar-looking' terms in their expressions.

A small number of candidates provided elegant alternative solutions, such as finding the area of triangle OBC and subtracting OAC, or (more commonly) finding CM and either AM or BM, then using 2x the area of triangle ACM or BCM.

Question 6 (b) (i)

(b) (i) Find the value of a when the area of triangle ABC is zero. **[1]**

This was generally well answered by the majority of candidates.

Question 6 (b) (ii)

(ii) Give a geometrical interpretation of the case in part **(b)(i)**. **[1]**

This (and part (c) below) was much less well answered. Many candidates attempted to compute either the area or radius (which was not required, and not sufficient). Relatively few candidates provided the required geometric explanation, with many making deductions like 'there are no solutions' or 'area is invalid', which were not acceptable.

A few candidates stated that A and B are coincident, without explaining that therefore $y = x$ is a tangent to the circle, which was acceptable.

Question 6 (c)

(c) Give a geometrical interpretation of the case where $a = 5$. **[1]**

As in part (b) (ii), many candidates gave an insufficient (or non-geometric) explanation. Lots of candidates simply stated that this case is 'impossible' or 'invalid' without explaining that this is because the line does not meet the circle.

Question 7 (a)

7 A student wishes to prove that, for all positive integers a and b, $a^2 - 4b \neq 2$.

(a) Prove that $a^2 - 4b = 2 \Rightarrow a$ is even. **[2]**

A range of methods were acceptable for both parts of this question, as proof by contradiction was not specified, although this was most commonly seen.

For this first part, candidates were able to make the required deduction fairly straightforwardly by rearranging the equation and explaining that the sum of two even numbers ($4b + 2$) must always be even, therefore a^2 is even, therefore so is a. However, many candidates attempted to prove the reverse implication by starting with 'a is even' (which was not acceptable). Some candidates attempted the alternative method by contradiction but this was generally less successful.

Candidates should be reminded that, although this is a 'Statistics' paper, the level of rigour required in proof questions remains consistent across the specification. In particular, it is important that candidates set up a contradiction correctly, stating a clear assumption and reaching a clear conclusion.

Some candidates provided much more algebraic working than was required in this part, although it was acceptable to simply state that the sum of two even numbers must be even (or to show this with $2(2b+2)$).

Question 7 (b)

(b) Hence or otherwise prove that, for all positive integers a and b, $a^2 - 4b \neq 2$. **[3]**

Many candidates did not realise that they could (although they were not required to) use the result from (a) to simplify the work in (b). Those who did were generally able to make good progress by writing $a = 2m$, substituting in and manipulating to reach an equation from which a contradiction can be shown.

Many candidates attempted to consider different cases for a and b each being odd/even (although only 2 of those cases needed to be considered following part (a)). In doing so, many used the same letter, e.g. writing $a = 2n$, $b = 2n + 1$ (which does not generalise so could be awarded the method mark but neither accuracy mark).

Some candidates spent time and working considering the cases where a is odd, although these were not relevant (because of part (a)) and therefore not credited.

In both parts of this question, many candidates did correct work but did not state a conclusion.

Section B overview

Most candidates coped well with the more routine questions in Section B (and it was not unusual for candidates to score better in Section B than in Section A). However, the less routine questions (Questions 11 and 14 in particular) were less well answered.

Question 8 (a)

8 The stem-and-leaf diagram shows the heights, in centimetres, of 15 plants.

0	2
1	0
2	4
3	0 2 4 9
4	1 2 4 7 9
5	3 7
6	2

Key: | 2 | 5 means 25 cm.

(a) Draw a box-and-whisker plot to illustrate the data. **[4]**

This was answered well by the majority of candidates, with some errors in calculating the median and quartiles, although relatively few candidates achieved full marks.

Most candidates were able to extract information from the stem-and-leaf diagram correctly, but some did not demonstrate the assumed knowledge in drawing a correct box and whisker plot – for instance, using the mean rather than median, or failing to mark the first and last values correctly.

Candidates needed to draw their own scale in this question, and while many selected the 'obvious' 1 square : 1 centimetre scale, some did not realise the need for a linear scale.

Some candidates decided to omit points they considered to be outliers. Although part (b) went on to ask about potential outliers (and in fact there were none using the IQR method) the question specifically states 'illustrate the data', so the full dataset needed to be represented on the diagram.

Question 8 (b)

A statistician intends to analyse the data, but wants to ignore any outliers before doing so.

(b) Discuss briefly whether there are any heights in the diagram which the statistician should ignore. **[3]**

Many candidates understood how to use either the quartiles or the mean and the standard deviation to find whether there were potential outliers (and, as the question did not specify, either method was acceptable here). However, candidates were required to support their response with calculations to gain full marks. Some candidates calculated correctly but lost the final mark by not stating whether or not there were outliers.

Of those who calculated using the mean and standard deviation (which required additional work because these had not been found in part (a)), many made errors in calculating or did not address that there are 'no (upper) outliers' after they had identified the lower outlier of 2.

Some candidates mixed up both methods, or subtracted the IQR from the median, which could not be credited.

Question 9 (a)

9 A school contains 500 students in years 7 to 11 and 250 students in years 12 and 13. A random sample of 20 students is selected to represent the school at a parents' evening. The number of students in the sample who are from years 12 and 13 is denoted by X.

(a) State a suitable binomial model for X. **[1]**

This was generally well answered, although some candidates erroneously gave n = 250 or 750, or gave an inexact value for p = 0.3 or 0.33. Some candidates gave p as 0.5 or 2/75.

Question 9 (b) (i)

Use your model to answer the following.

(b) (i) Write down an expression for $\mathrm{P}(X = x)$. **[1]**

Many candidates were able to give this expression for the standard Binomial probability correctly (often with follow-through using their values from part (a)). Some candidates omitted the combination term or gave the general expression without values substituted in, which was not enough for the mark.

Question 9 (b) (ii)

(ii) State, in set notation, the values of x for which your expression is valid. **[1]**

Very few candidates obtained this mark. Some identified the correct range of values, in correct set notation, but hardly any candidates stated that only integer values in the range were appropriate.

Question 9 (c)

(c) Find $\mathrm{P}(5 \leqslant X \leqslant 9)$. **[2]**

Candidates' responses to this question were very mixed, with only a minority of candidates obtaining the correct answer. Many candidates did not appear to recognise which probabilities they needed to calculate and subtract, or used a complement such as $1 - P(X \leq 4)$.

The method mark was awarded for inaccurate working such as $P(X \leq 9) - P(X \leq 5)$, which was a common answer (giving 0.611).

Question 9 (d)

(d) State one disadvantage of using a random sample in this context. **[1]**

A good number of candidates answered this correctly, providing a sufficient explanation in context (acknowledging the idea that with a small sample it may not provide proportionate representation across the year groups).

Some candidates gave non-mathematical disadvantages, such as the suitability of particular individuals to represent the school (which was not an acceptable answer), or simply stated the generic fact that 'the sample would not be representative' or 'a stratified sample would be better' without putting this in context – such as referring to the year groups.

Question 10

10 The mass, in kilograms, of a species of fish in the UK has population mean 4.2 and standard deviation 0.25.

An environmentalist believes that the fish in a particular river are smaller, on average, than those in other rivers in the UK.

A random sample of 100 fish of this species, taken from the river, has sample mean 4.16 kg.

Stating a necessary assumption, test at the 5% significance level whether the environmentalist is correct. **[8]**

The numerical part of this question was well answered by the majority of candidates, with most demonstrating sound knowledge of the core technique for hypothesis testing.

Very few candidates achieved the mark for stating the necessary assumption (which needed to be specific to the fish in this river, recognising that the standard deviation for the whole of the UK is given in the question), even though they made use of this in their answer.

Some candidates did not achieve both marks for the 'setup' of the hypothesis test, mainly because of failing to define their μ correctly, or not providing a definition at all.

The remaining 5 marks for carrying out the test were very often obtained in full, with most candidates choosing to compare probabilities. A small number used critical values. There were some errors in stating the final conclusion, with a small number of candidates either providing a 'definite' conclusion or mis-stating what the test shows, for instance wrongly asserting that the mean has 'not decreased'. The context here did not provide any reference to change over time, so candidates should be careful to make sure that their response is specific to the context of a given scenario.

Question 11 (a)

11 The random variable Y has the distribution $\mathrm{N}(\mu, \sigma^2)$.

(a) Find $\mathrm{P}(Y > \mu - \sigma)$. **[1]**

This was generally well answered, with many candidates calculating the required value exactly (to get 0.841). Some candidates used the 68% 'rule of thumb' approximation, which was acceptable provided it was applied correctly (obtaining 0.84). Using the less accurate 2/3 'rule of thumb' approximation, and then further approximating this to 66%, was not deemed sufficiently accurate given that the question asked candidates to 'Find'.

Some candidates stated the probability only as, e.g. 0.16 or 0.34 only, without undertaking the required subtraction or division by 2.

Question 11 (b)

(b) Given that $P(Y > 45) = 0.2$ and $P(Y < 25) = 0.3$, determine the values of μ and σ. **[6]**

Many candidates achieved full marks on this question, although a reasonable number of candidates appeared not to understand the request and so made very little progress.

Most candidates were able to setup the two required equations (with a few making a sign error or using 0.8 and 0.3 instead of the inverse normal values), and these were generally then solved correctly to the required degree of accuracy (3sf).

The command word 'Determine' here indicates that sufficient working was required to support the answer, but it does not mean that candidates are forbidden from using a calculator – as many did – to solve the simultaneous equations, provided that their solution can be followed from a clearly set out method using the information given in the question.

Question 11 (c)

The random variables U and V have the distributions $N(10, 4)$ and $N(12, 9)$ respectively.

(c) It is given that $P(U < b) = P(V > c)$, where $b > 10$ and $c < 12$.

Determine b in terms of c. **[2]**

Very few candidates obtained any marks in this question, and many candidates did not make much of an attempt. Of those who did, a good number went on to obtain the correct solution (in general, candidates who wrote down the correct equation for the M mark went on to solve correctly and obtain the A mark). Some candidates omitted the '−' sign or wrote down the complement of one expression (1−…), obtaining an incorrect equation in *b* and *c*.

Question 12 (a)

12 A student has an ordinary six-sided dice. The student suspects that it is biased against six, so that when it is thrown, it is less likely to show a six than if it were fair.

In order to test this suspicion, the student plans to carry out a hypothesis test at the 5% significance level.

The student throws the dice 100 times and notes the number of times, X, that it shows a six.

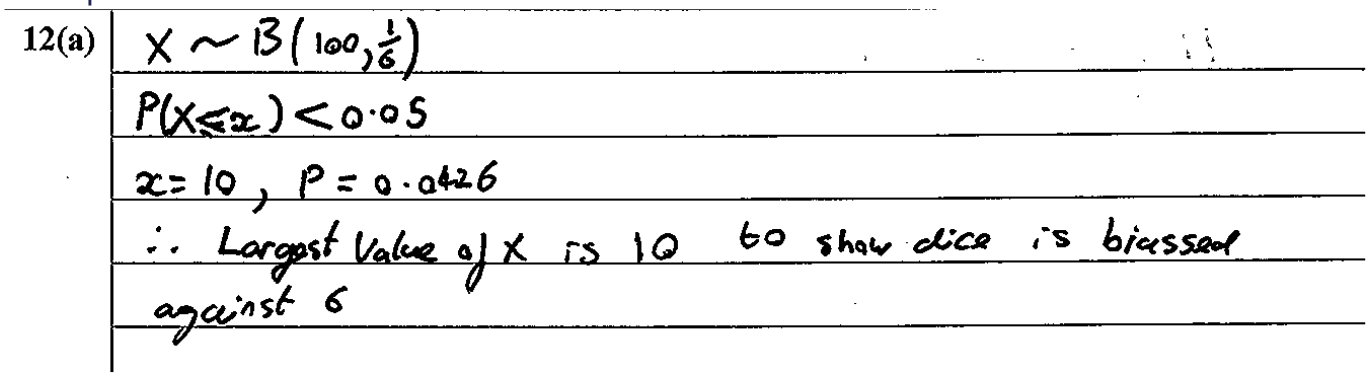
(a) Determine the largest value of X that would provide evidence at the 5% significance level that the dice is biased against six. **[3]**

This question was broadly well answered, with many candidates successfully concluding with the correct answer of 10.

Some candidates did not show sufficient working (for 'Determine') to support their conclusion, or were not sufficiently accurate with the probabilities they were using to support their answer (these could be rounded, but needed to be rounded correctly or '…' used to show truncation).

A small number of candidates attempted to use a normal approximation, which was a good (if laborious) way of identifying the borderline values to test, but then did not go on to calculate the Binomial probabilities required to support a final conclusion. Some candidates used the pdf rather than the cdf (obtaining incorrect values).

Exemplar 3



This candidate has found one of the required probabilities (although they have 'chopped' their answer so it is not rounded correctly), but has then concluded directly – without providing the value for $x = 11$ to justify this as the largest value, so they could only be awarded the first M mark.

Question 12 (b)

Later another student carries out a similar test, at the 5% significance level. This student also throws the dice 100 times.

(b) It is given that the dice is fair.

Find the probability that the conclusion of the test is that there is significant evidence that the dice is biased against six. **[1]**

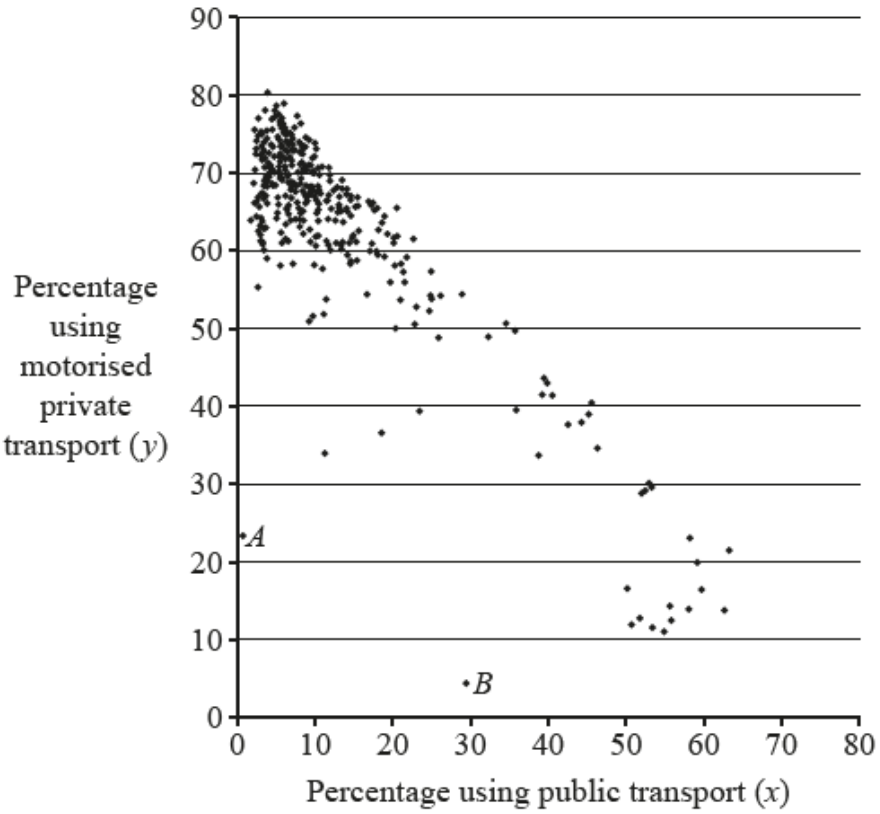
This question was only correctly answered by a minority of candidates. Many candidates appeared to misunderstand that the critical region for a test (with a discrete distribution) is not exactly 0.05, and gave this as an incorrect answer (even though they may have calculated the required probability in part (a)).

Question 13 (a) (i)

13 The scatter diagram uses information about all the Local Authorities (LAs) in the UK, taken from the 2011 census.

For each LA it shows the percentage (x) of employees who used public transport to travel to work and the percentage (y) who used motorised private transport.

"Public transport" includes train, bus, minibus, coach, underground, metro and light rail.
"Motorised private transport" includes car, van, motorcycle, scooter, moped, taxi and passenger in a car or van.



(a) Most of the points in the diagram lie on or near the line with equation $x+y=k$, where k is a constant.

(i) Give a possible value for k. **[1]**

The majority of candidates answered this correctly, giving a plausible value for k between 70 and 80. Some candidates over-estimated (giving a not-plausible value of, e.g. 85 or more) and a few candidates attempted to find the gradient of the line.

Question 13 (a) (ii)

(ii) Hence give an approximate value for the percentage of employees who either worked from home or walked or cycled to work. **[1]**

Many candidates correctly computed 100 – (their *k*) here, although quite a few did not heed the 'Hence' instruction and gave a value which was not compatible with their response to (a)(i). This was an important point of understanding (and one that was relied on later in the question) for candidates to recognise that there are effectively three categories being used in this question: 'motorised private transport', 'public transport' and 'walk/cycle/work from home' and that together they must sum to 100% for a given LA.

Question 13 (b) (i)

(b) The average amount of fuel used per person per day for travelling to work in any LA is denoted by F.

Consider the two groups of LAs where the percentages using motorised private transport are highest and lowest.

(i) Using only the information in the diagram, suggest, with a reason, which of these two groups will have greater values of F than the other group. **[1]**

This question required candidates to clearly ('unambiguously' in the mark scheme) identify the group of LAs which has greater values of F, and then provide a reason for this.

Some candidates misinterpreted and referred to A or B (which are individual LAs, not within the groups being referred to here). Many candidates did not identify the group sufficiently clearly, making statements like 'those who use public transport' (noting that neither percentage is 0 for any LA in either group referred to, so this does not unambiguously identify the group).

Some candidates omitted a reason, but those who included one generally gave a sensible reason which justified their response in context. It is important here that the reason required was one which explains why this group has a higher value of F (for instance because public transport may have more passengers per vehicle than private transport), not just identifying where the group is found on the scatter diagram.

Question 13 (b) (ii)

A student says that it is not possible to give a reliable answer to part **(b)(i)** without some further information.

(ii) Suggest two kinds of further information which would enable a more reliable answer to be given. **[2]**

Many candidates answered this well, giving two sensible suggestions. A broad range of answers were seen.

Some candidates suggested that the population sizes of the LAs would improve the reliability of the answer, which is not correct because F is defined as fuel used per person.

A few candidates gave two suggestions which were effectively duplicates of each other, so only received 1 mark.

Question 13 (c) (i)

(c) Points A and B in the diagram are the most extreme outliers. Use their positions on the diagram to answer the following questions about the two LAs represented by these two points.

(i) The two LAs share a certain characteristic.

Describe, with a justification, this characteristic. **[2]**

Many candidates struggled with one or both parts of (c). Here, for the 2 marks to be given candidates needed to clearly describe the shared characteristic, and provide a justification for it.

The shared characteristic (which might be thought of as 'how could these two LAs be uniquely identified within the data set') needed to specify that these two LAs have the lowest **combined** usage of both types of transport (private and public) or that they had the highest percentages of employees walking/cycling/working from home. The key observation is that A and B are (in effect) the furthest 'bottom left' outliers from the line of best fit, so a description in terms of the horizontal or vertical axis alone does not uniquely identify them.

For the justification, again candidates needed to provide a rationale **in context** as to why this might be the case. The simplest acceptable answer was simply that these LAs may be the smallest (in terms of their physical area – which turns out to be correct as they are the Isles of Scilly and the City of London). Many candidates went on to give a more detailed explanation (e.g. by quoting values from the diagram) without providing this justification in context.

Question 13 (c) (ii)

(ii) The environments in these two LAs are very different.

Describe, with a justification, this difference. **[2]**

Many candidates appeared to find this question challenging, but those who correctly interpreted the careful language in the question were able to provide a fully correct response. As in part (i) candidates needed to clearly state the difference in environments, and then provide a justification in context.

Some candidates were able to use their knowledge of the Large Data Set (and the presumption that these two LAs are often outliers) to help with their answer, but this was not required.

Some candidates correctly identified that one LA was likely to be rural, and the other urban, but did not identify which was which – so could gain no credit. The statement of difference needed to be at least comparative (so it was acceptable to say 'A is likely to be more rural than B' but not just 'A is rural').

For the justification, candidates needed to explain **why** this difference might arise – rather than just re-stating the data. So, for instance, candidates who provided acceptable responses made statements to the effect that there may be less availability of public transport in A, or factors that limit the use of private transport in B. Statements that repeated the data (e.g. 'A has lower public transport use') were not sufficient alone.

Some candidates made general statements about environmental issues, which could not be credited.

Question 13 (d)

(d) A student says that it is difficult to extract detailed information from the scatter diagram.

Explain whether you agree with this criticism. **[1]**

Many candidates answered this question well, but quite a number did not identify that the specific request here is for a critique of the diagram – and why this particular scatter diagram may make it difficult to extract detailed information.

Some candidates highlighted additional information which may have been useful – which is not what this question required. Acceptable responses gave a specific reason as to why this diagram (i.e. this specific presentation of the data) may limit the extraction of detailed information – for instance, that the points are too closely clustered together, or that there are too few gridlines to read values precisely.

Explanations

When asked for explanations, candidates should carefully examine the question to make sure they understand the request. Explanations should always be provided in context. A justification needs to (in statistics) explain **why** a particular pattern or characteristic in the data may be observed, not just describe it further.

Question 14 (a)

14 In this question you must show detailed reasoning.

A disease that affects trees shows no visible evidence for the first few years after the tree is infected.

A test has been developed to determine whether a particular tree has the disease. A positive result to the test suggests that the tree has the disease. However, the test is not 100% reliable, and a researcher uses the following model.

- If the tree has the disease, the probability of a positive result is 0.95.
- If the tree does not have the disease, the probability of a positive result is 0.1.

(a) It is known that in a certain county, A, 35% of the trees have the disease. A tree in county A is chosen at random and is tested.

Given that the result is positive, determine the probability that this tree has the disease. **[3]**

A substantial minority of candidates answered this question correctly, although with it being the final question on the paper it may be that some were running short of time.

Those who made an attempt were likely to obtain the correct answer, demonstrating their knowledge of conditional probability to obtain the required answer.

Some candidates did not use the conditional probability formula and just multiplied two values together.

Question 14 (b)

A forestry company wants to determine what proportion of trees in another county, B, have the disease. They choose a large random sample of trees in county B.

Each tree in the sample is tested and it is found that the result is positive for 43% of these trees.

(b) By carrying out a calculation, determine an estimate of the proportion of trees in county B that have the disease. **[4]**

Very few candidates achieved 3 or 4 marks in this part. Some misinterpreted the question and used 0.43 as a further probability (e.g. multiplying it by 0.1 or 0.43) rather than setting up the required equation in terms of p (or other variable).

Of those who did the question correctly, the majority used the solution set out in the mark scheme. A minority wrote $0.95a + 0.1b = 0.43$, $0.05a + 0.9b = 0.57$ and then used simultaneous equations to find the answer correctly.

Almost all candidates who correctly found the answer to be 0.388 did not go on to (as was required) state an approximation in context, and so could not be awarded the final B mark. For this mark, candidates needed to state their answer in context, and show clearly that they understood it was an approximation – for instance by using the word 'approximately'.

Supporting you

Teach Cambridge

Make sure you visit our secure website **Teach Cambridge** to find the full range of resources and support for the subjects you teach. This includes secure materials such as set assignments and exemplars, online and on-demand training.

Don't have access? If your school or college teaches any OCR qualifications, please contact your exams officer. You can forward them this link to help get you started.

Reviews of marking

If any of your students' results are not as expected, you may wish to consider one of our post-results services. For full information about the options available visit the OCR website.

Access to Scripts

For the June 2023 series, Exams Officers will be able to download copies of your candidates' completed papers or 'scripts' for all of our General Qualifications including Entry Level, GCSE and AS/A Level. Your centre can use these scripts to decide whether to request a review of marking and to support teaching and learning.

Our free, on-demand service, Access to Scripts is available via our single sign-on service, My Cambridge. Step-by-step instructions are on our website.

Keep up-to-date

We send a monthly bulletin to tell you about important updates. You can also sign up for your subject specific updates. If you haven't already, sign up here.

OCR Professional Development

Attend one of our popular CPD courses to hear directly from a senior assessor or drop in to a Q&A session. Most of our courses are delivered live via an online platform, so you can attend from any location.

Please find details for all our courses for your subject on **Teach Cambridge**. You'll also find links to our online courses on NEA marking and support.

Signed up for ExamBuilder?

ExamBuilder is the question builder platform for a range of our GCSE, A Level, Cambridge Nationals and Cambridge Technicals qualifications. Find out more.

ExamBuilder is **free for all OCR centres** with an Interchange account and gives you unlimited users per centre. We need an Interchange username to validate the identity of your centre's first user account for ExamBuilder.

If you do not have an Interchange account please contact your centre administrator (usually the Exams Officer) to request a username, or nominate an existing Interchange user in your department.

Active Results

Review students' exam performance with our free online results analysis tool. It is available for all GCSEs, AS and A Levels and Cambridge Nationals.

Find out more.

Need to get in touch?

If you ever have any questions about OCR qualifications or services (including administration, logistics and teaching) please feel free to get in touch with our customer support centre.

Call us on
01223 553998

Alternatively, you can email us on
support@ocr.org.uk

For more information visit
ocr.org.uk/qualifications/resource-finder
ocr.org.uk
facebook.com/ocrexams
twitter.com/ocrexams
instagram.com/ocrexaminations
linkedin.com/company/ocr
youtube.com/ocrexams

We really value your feedback

Click to send us an autogenerated email about this resource. Add comments if you want to. Let us know how we can improve this resource or what else you need. Your email address will not be used or shared for any marketing purposes.

I like this **I dislike this**

Please note – web links are correct at date of publication but other websites may change over time. If you have any problems with a link you may want to navigate to that organisation's website for a direct search.

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT