Qualification Accredited

OCR
Oxford Cambridge and RSA

A LEVEL

Examiners' report

MATHEMATICS A

H240

For first teaching in 2017

H240/02 Summer 2022 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.

Advance Information for Summer 2022 assessments

To support student revision, advance information was published about the focus of exams for Summer 2022 assessments. Advance information was available for most GCSE, AS and A Level subjects, Core Maths, FSMQ, and Cambridge Nationals Information Technologies. You can find more information on our website.

Paper 2 series overview

The overall standard was similar to that in 2019 (i.e. the last relatively "normal" year), with a range of marks from very good to fairly low. Very few candidates scored below 20%. Responses to the questions requiring verbal answers (3 (a), 9 (b) (i) and (ii), 9 (c) (ii),10 (a) (i), (ii) and (iii), 10 (c), 10 (d), 10 (e) and 12 (b)) were variable, with many candidates not answering precisely the question that was set. In many cases candidates wrote far more than was necessary to gain the mark(s). Centres are referred to the published mark scheme to see the kind of verbal answers that are acceptable.

Some candidates resorted to trial and improvement methods. These are usually not given full credit, even if the correct answers are obtained. (Although, unusually, Question 4 was an exception this year.)

Some candidates did not show sufficient working, and consequently lost marks. Candidates need to be made aware of the significance of words such as "detailed reasoning" and "determine", as explained below.

Answering the actual question that is asked

Perhaps more so than in other years, there were many examples of candidates not answering the actual question before them. In some cases, a fairly good response was given, but it was a response to a slightly different question from the one that was asked. See especially the comments on Questions 2 (a), 3 (b) (i), 3 (c), 4, 9 (e) (ii), 10 (a), 10 (d), 10 (e) and 12 (b).

Handwriting

Many candidates' scripts were difficult to read. This problem seemed worse than in past years, and needs to be addressed by centres.

How much working needs to be shown to gain full marks?

Some candidates seemed unsure as to how much working was required for any given question. This problem arises from what might seem a dilemma. On the one hand, candidates are expected to be able to use calculator functions that "short cut" techniques such as definite integration, solution of quadratic and cubic equations, mean and standard deviation. On the other hand, candidates are also expected to show understanding of precisely the techniques that these functions make unnecessary. This dilemma is resolved by the careful use, in questions, of certain command words, whose definitions are found on pages 9, 10 and 11 of the specification. This can be illustrated by comparing Question 9 (d) with Questions 1 (b) and 5 (a) (which both contain the instruction **In this question you must show detailed reasoning**).

In Question 1 the instruction means that, when solving the quadratic equation that arises in the solution of part (b), candidates are expected to show a method (either factorisation, or the formula, or completing the square) for obtaining the solutions. Candidates who obtained the quadratic equation and then just wrote down the correct solutions lost 1 method mark. Similarly, in Question 5 (a), many candidates found the correct quadratic equation, and then just stated that it had no real roots, without showing a method for obtaining this conclusion. These candidates lost a mark. In contrast to this, Question 9 (d) uses the word "find". This is not one of the command words defined in the specification, and so there is no requirement to show all the working. It was expected that, having found the frequencies, candidates would use the statistical functions on the calculator to write down the mean and standard deviation without explanation. Many candidates did this successfully. Many other candidates, however, showed working and in some cases made arithmetical errors and lost the relevant marks. Of course, candidates are free to show working if they wish to.

When using one of the "short cut" functions in questions like 9 (d), candidates would be well advised to carry out the technique twice as a check.

The command word "Determine" also implies that each stage of the working must be shown. Thus, for example, in Question 12 (a) candidates who gave the correct answer without working, or with inadequate working, scored only 3 marks out of 5.

It is worth noting that in Question 11, since there is no instruction about detailed reasoning, nor a "Determine", the calculation required in this question can be performed without showing working. For example, having stated the distribution of $\overline{X}$, it is sufficient to write

$P(\overline{X} > 3360) = 0.0297$, or $P(\overline{X} > a) = 0.025 \Rightarrow a = 3362$.

There is no need to show working involving $\dfrac{\overline{X}-3300}{450/\sqrt{200}}$.

Candidates who did well on this paper generally did the following:	Candidates who did less well on this paper generally did the following:
• demonstrated good algebraic skills • were able to interpret a statistical table and a statistical diagram • were able to break a problem down into its constituent parts • understood when to use calculator functions and when to show detailed working • were able to make a good attempt at a complex probability question • had a good understanding of the concept of proof.	• demonstrated limited algebraic skills • were unable to solve quadratic equations • did not have a good understanding of hypothesis testing • showed no understanding of differential equations • were able to attempt only simple probability questions • did not have a good understanding of the concept of proof • did not recall the appropriate methods for integration.

Section A overview

Candidates generally answered Questions 1, 2 (a) and 5 (a) well although, for many, their algebraic skills let them down. Manipulation of vectors in Question 2 (b) was generally not very successful, and work on the differential equation (Question 8) was on the whole not good.

Question 1 (a)

1 In this question you must show detailed reasoning.

Solve the following equations.

(a) $\frac{x}{x+1} - \frac{x-1}{x+2} = 0$ **[3]**

Most candidates successfully either collected over a common denominator or "cross-multiplied". A few made a sign error as follows:

$x^2 + 2x - (x - 1)(x + 1) = x^2 + 2x - x^2 - 1$. Some equated the denominator to zero and obtained extra, bogus answers.

Question 1 (b)

(b) $\frac{8}{x^6} - \frac{7}{x^3} - 1 = 0$ **[3]**

Most candidates formed a quadratic equation, either in x^3 or in $\frac{1}{x^3}$. Some then used a substitution, while others solved the equation as it was. Most were successful. A few candidates stated that $x^3 = -8$ gave no solution. Many candidates lost a mark because they did not show a method for obtaining the roots of the quadratic equation. Some multiplied the equation by x^9 throughout, and eventually obtained an extra (incorrect) answer of $x = 0$.

A minority of candidates could not solve the quadratic equation, giving working such as

$\frac{8}{x^6} - \frac{7}{x^5} - 1 = 0 \Rightarrow \frac{8}{x^6} - \frac{7}{x^5} - 1 = 0 \Rightarrow 8 - 1 = \frac{7x^6}{x^3} \Rightarrow 7 = 7x^3 \Rightarrow x = 1$, or

$8 - 7x^3 = x^9 \Rightarrow 8 = x^9 - 7x^3 \Rightarrow 8 = x^3(x^3 - 7) \Rightarrow x^3 = 7 \Rightarrow x = \sqrt[3]{7}$.

Question 1 (c)

(c) $3^{x^2-7} = \frac{1}{243}$ **[2]**

This question was well answered by most candidates. A wide variety of methods was seen, some involving logs, others involving manipulation of indices. Some candidates omitted the ± in their answer. A few candidates showed lacking understanding of logs, such as $x^2 - 7 = \ln\frac{1}{243}$, or of indices, such as $\frac{1}{3^7}x^2 = \frac{1}{243}$.

Question 2 (a)

2 The points A and B have position vectors $3\mathbf{i} + 2\mathbf{j}$ and $4\mathbf{i} + 2\mathbf{j} - 5\mathbf{k}$ respectively.

(a) Find the length of AB. **[2]**

This question was answered well by most candidates. A few found $\overrightarrow{AB}$, but did not go on to find the length, perhaps because of a misreading of the question. Some others thought that $\overrightarrow{AB}$ = **a** + **b**. Some did not use good notation, omitting brackets: $1^2 + -5^2 = 26$.

Question 2 (b)

Point P has position vector $p\mathbf{i} - 3\mathbf{k}$, where p is a constant. P lies on the circumference of a circle of which AB is a diameter.

(b) Find the two possible values of p. **[3]**

Perhaps the quickest way to approach this question is to use Pythagoras' theorem. However, very few candidates used this approach. Many attempted to find the position vector of the centre, C, of the circle and then used $|PC|^2$ = radius^2. These candidates generally scored the first mark, but lost the other marks by making arithmetical slips. Some omitted the **j** component. Some omitted to halve the diameter. A common, more serious, error was $|OP|^2$ = radius^2.

Question 3 (a)

3 **(a)** Amaya and Ben integrated $(1+x)^2$, with respect to x, using different methods, as follows.

Amaya: $\int (1+x)^2\,dx = \dfrac{(1+x)^3}{3} + c \qquad = \frac{1}{3} + x + x^2 + \frac{1}{3}x^3 + c$

Ben: $\int (1+x)^2\,dx = \int (1+2x+x^2)\,dx \qquad = x + x^2 + \frac{1}{3}x^3 + c$

Charlie said that, because these answers are different, at least one of them must be wrong.

Explain whether you agree with Charlie's statement. **[1]**

Many candidates saw the point and made reasonable comments about the arbitrary constant. Some wrote that both methods were correct, with no further explanation. These responses did not score the mark. Other candidates stated that Amaya's (or Ben's) method was incorrect.

Question 3 (b) (i)

(b) You are given that a is a constant greater than 1.

(i) Find $\int_1^a \dfrac{1}{(1+x)^2}\,dx$, giving your answer as a single fraction in terms of the constant a. **[3]**

This question was answered well by many candidates. Some correctly obtained $\frac{1}{2} - \frac{1}{1+a}$, but either omitted the simplification (presumably not having read the question carefully) or made an algebraic error when attempting the simplification. A small number of candidates substituted the limits the wrong way round. Others substituted the limits, but then added instead of subtracting. Many candidates gave incorrect answers involving either $(1 + a)^{-3}$ or $\ln(1 + x)$, and a few candidates attempted to use partial fractions (of the form $\frac{A}{(1+x)^2} + \frac{B}{(1+x)}$) before integrating.

Question 3 (b) (ii)

(ii) You are given that the area enclosed by the curve $y = \frac{1}{(1+x)^2}$, the x-axis and the lines $x = 1$ and $x = a$ is equal to $\frac{1}{3}$.

Determine the value of a. **[2]**

Almost all candidates recognised that they could simply equate their answer to part (b) (i) to $\frac{1}{3}$, although a few started from scratch by integrating again. Some candidates who had impossibly difficult incorrect answers to part (b) (i), gave up the attempt to solve their equation, and just wrote down the correct answer of *a* = 5. Presumably these candidates had used the integration function on their calculator, and had tried different values of *a* until they found the one that gave an answer of $\frac{1}{3}$. Because this question is a "Determine" these answers were not accepted.

Question 3 (c)

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

Find the exact value of $\int_0^{\frac{1}{12}\pi} \frac{\cos 2x}{\sin 2x + 2}\,\mathrm{d}x$, giving your answer in its simplest form. **[4]**

There were many neat and correct solutions. Many candidates used a substitution. However, many others used the more efficient method that comes from recognising that the numerator is *k* × the derivative of the denominator. Some candidates omitted "*k*" or had *k* = 2 instead of $\frac{1}{2}$. A few candidates correctly obtained $\frac{1}{2}\left(\ln(\frac{5}{2}) - \ln 2\right)$, but omitted the final step of simplification (presumably, as in part (a) (i), because they had not read the question carefully). Others used decimals, ignoring the word "exact" in the question.

There were also some unsuccessful attempts at this question. Some candidates used the double angle formulae. Others attempted to use integration by parts on cos 2*x* × (sin 2*x* + 2)$^{-1}$. A few rewrote the integrand as $\frac{\cos 2x}{\sin 2x} + \frac{\cos 2x}{2}$. None of these attempts yielded any marks.

Question 4

4 An artist is creating a design for a large painting. The design includes a set of steps of varying heights. In the painting the lowest step has height 20 cm and the height of each other step is 5% less than the height of the step immediately below it.

In the painting the total height of the steps is 205 cm, correct to the nearest centimetre.

Determine the number of steps in the design. **[5]**

Almost all candidates recognised the need to consider the sum of a geometric progression. (Although a few used either the formula for the n^{th} term of a GP or the formula for the sum of n terms of an AP.) In most cases, there was some good algebraic manipulation, including the use of logs to solve an equation such as $0.95^n = \frac{39}{80}$. There were two common errors. Firstly some candidates used $r = 1.05$ (representing increasing, rather than decreasing, heights). Others used $r = \frac{1}{0.95}$ (resulting from a misreading of the word "lowest" in the question to mean "shortest"). Candidates who made either of these errors, but whose working was otherwise correct, scored 4 marks out of 5. A more serious error was to use $r = 0.05$. Candidates who made this error could only score one mark for recognising a geometric progression.

Question 5 (a)

5 **In this question you must show detailed reasoning.**

A curve has equation $y = x^3 - 3x^2 + 4x$.

(a) Show that the curve has no stationary points. **[2]**

Almost all candidates differentiated and equated $\frac{dy}{dx}$ to 0. A few made errors in the differentiation. Many obtained the correct equation, $3x^2 - 6x + 4 = 0$. Because this is a "detailed reasoning" question, it was essential to show working to demonstrate that this equation has no real roots in order to gain the second mark. Most did so successfully using either the discriminant or completing the square. However, some candidates just stated that there were no real solutions, without justification. A few gave an incorrect justification such as "this equation won't factorise so there are no solutions". More seriously, a few candidates used the coefficients from the cubic expression for y to obtain a bogus "discriminant" and showed that it was negative.

Question 5 (b)

(b) Show that the curve has exactly one point of inflection. **[2]**

Most candidates found $\frac{d^2y}{dx^2}$, equated it to 0, and correctly obtained $x = 1$. In order to gain the second mark, it was necessary to comment that this meant there was an inflection here, or that there was only one root. This was enough to answer the question, although some candidates felt they needed to go further and show that $x = 1$ gives an inflection, by considering the sign of $\frac{dy}{dx}$ or $\frac{d^2y}{dx^2}$ either side. This was unnecessary, since the two conditions $\frac{dy}{dx} \neq 0$ and $\frac{d^2y}{dx^2} = 0$ have both been shown to be satisfied. Some candidates found the y-coordinate at $x = 1$, which was also unnecessary.

Question 6 (a)

6 (a) The diagrams show five different graphs. In each case the whole of the graph is shown.

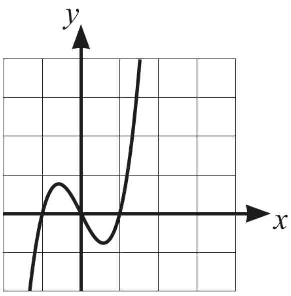


Fig. 1.1

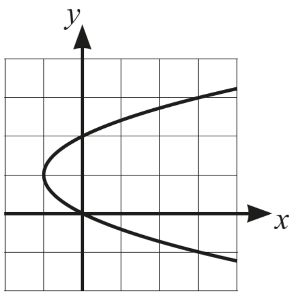


Fig. 1.2

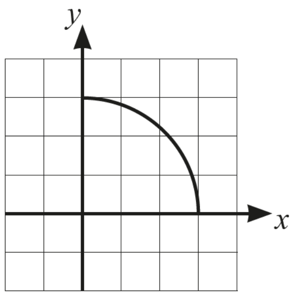


Fig. 1.3

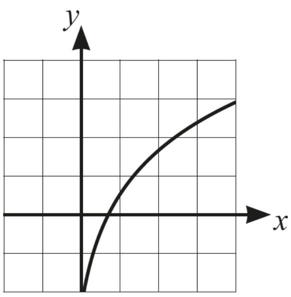


Fig. 1.4

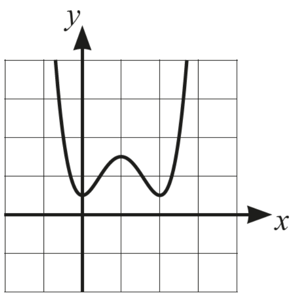


Fig. 1.5

Place ticks in the boxes in the table in the Printed Answer Booklet to indicate, for each graph, whether it represents a one-one function, a many-one function, a function that is its own inverse or it does not represent a function. There may be more than one tick in some rows or columns of the table. **[4]**

Responses to this question were mixed. The responses for diagrams 1.4 and 1.5 were usually correct. Perhaps the most difficult aspect of the question was identifying the curve that represented a self-inverse function.

Question 6 (b)

(b) A function f is defined by $f(x) = \frac{1}{x}$ for the domain $\{x : 0 < x \leqslant 2\}$.

State the range of f, giving your answer in set notation. **[2]**

Many candidates identified the range correctly, although not all could express it correctly in set notation. Some common answers that gained only 1 mark were these $y \geq \frac{1}{2}$, $\{x: x \geq \frac{1}{2}\}$, $\{f(x): f(x) > \frac{1}{2}\}$, $\{x: f(x) \geq \frac{1}{2}\}$, $(\frac{1}{2}, \infty)$. Some unnecessary "∞"s were seen (though not penalised), e.g. $\{y: \frac{1}{2} \leq y \leq \infty\}$.

Some common wrong answers involved $f(x) \leq \frac{1}{2}$ or $f(x) > 2$.

Question 7 (a)

7 It is given that any integer can be expressed in the form $3m+r$, where m is an integer and r is 0, 1 or 2.

Use this fact to answer the following.

(a) By considering the different values of r, prove that the square of any integer **cannot** be expressed in the form $3n+2$, where n is an integer. **[4]**

The concept of proof seemed to be better understood than in the past, with many candidates giving a fully correct response. There are many different correct approaches to this question, all of which were accepted. Four such approaches are given in the published mark scheme, although this list is not exhaustive.

A few candidates used what is, perhaps, the most elegant method which involved writing $(3m + r)^2$ in the form $3(3m^2 + 2mr) + r^2$, and then noting that $r^2 = 0$, 1 or 4 and so cannot be 2.

Some candidates showed numerical examples of three consecutive integers whose squares cannot be expressed in the form $3n + 2$. Some claimed (incorrectly) that this proved the desired result "by counter example".

Perhaps the most common approach was to square $3m$, $3m + 1$ and $3m + 2$. This alone would gain 2 marks. But many candidates did not gain the remaining marks because they proceeded simply to state, without justification, that none of these expansions could be expressed in the form $3n + 2$. To gain the last 2 marks, the expansions of $(3m + 1)^2$ and $(3m + 2)^2$ had to be written in the forms $3(3m^2 + 2m) + 1$ and $3(9m^2 + 4m + 1) + 1$ or similar.

Some candidates began by considering odd and even integers separately. Although this approach has been fruitful for answering some questions in previous years, it was not helpful in this case.

Question 7 (b)

(b) Three integers are chosen at random from the integers 1 to 99 inclusive. The three integers are not necessarily different.

By considering the different values of r, determine the probability that the sum of these three integers is divisible by 3. **[4]**

This rather unusual question puzzled many candidates. However, many made a reasonable attempt at it, using native wit rather than learned techniques. Those who used an algebraic approach, with $3m$, $3m + 1$ and $3m + 2$ made some progress in recognising the different cases (three integers being either all of the same type or of three different types). But they usually did not obtain the correct number of combinations that are possible. Those who actually listed the possibilities were more likely to be successful.

However, even this method was not always pursued successfully. In some cases, some of the possibilities were omitted. One particular example was as follows.

The possible combinations of values of r are these: 000, 111, 222, 012, 112, 221, 110, 001, 220, 002.

Four of these 10 combinations have a total that is divisible by 3, hence P(total is divisible by 3) = $\frac{4}{10}$.

Another incorrect approach was this: P(an integer is divisible by 3) = $\frac{33}{99}$, hence

P(the total is divisible by 3) = $\left(\frac{33}{99}\right)^3 = \frac{1}{27}$.

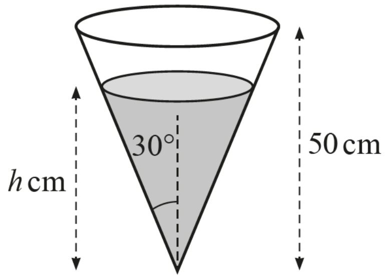
An interesting incorrect method that leads to the correct answer was as follows.

For the sum to be divisible by 3, all three integers must be of the form $3m$ OR all three must be either of the form $3m + 1$ or $3m + 2$.

Hence P(sum is divisible by 3) = $[P(r = 0)]^3 + [1 - P(r = 0)]^3 = \left(\frac{1}{3}\right)^3 + \left(\frac{2}{3}\right)^3 = \frac{1}{27} + \frac{8}{27} = \frac{1}{3}$.

Question 8

8



The diagram shows a water tank which is shaped as an inverted cone with semi-vertical angle 30° and height 50 cm. Initially the tank is full, and the depth of the water is 50 cm.

Water flows out of a small hole at the bottom of the tank. The rate at which the water flows out is modelled by $\frac{dV}{dt} = -2h$, where V cm^3 is the volume of water remaining and h cm is the depth of water in the tank t seconds after the water begins to flow out.

Determine the time taken for the tank to become empty.

[For a cone with base radius r and height h the volume V is given by $\frac{1}{3}\pi r^2 h$.] **[7]**

This differential equation question is of a fairly standard type, found in many text books. So one might reasonably expect candidates to be familiar with questions of this type. Despite this, responses were, on the whole, disappointing. In fact, about half the candidates scored 0 marks on this question.

However, there were also a good number of excellent solutions.

A few candidates thought that $r = h\sin 30°$, but gave an otherwise completely correct solution. These candidates scored 5 marks out of 7.

The main difficulty faced by candidates was that there were more than two variables involved (V, h, t and r). Most candidates obtained an equation in three variables, but then attempted to differentiate or integrate, treating one of the variables as a constant. For example:

$V = \frac{1}{3}\pi r^2 h \Rightarrow \frac{dV}{dr} = \frac{2}{3}\pi rh$ (or $\frac{dV}{dh} = \frac{1}{3}\pi r^2$).

$\frac{dV}{dh} = -2t \Rightarrow dV = -2t\,dh \Rightarrow V = -2th$.

Some candidates avoided this trap by using a numerical value for one of the variables, e.g. $h = 50$ or $r = \frac{50\sqrt{3}}{3}$, thus rendering their equation incorrect.

Some candidates appreciated the need to use the chain rule but, while doing so, made one of the above errors. Thus a typical solution was as follows. $\frac{dV}{dh} = \frac{dV}{dt} \times \frac{dt}{dh}$, but $\frac{dV}{dh} = \frac{1}{3}\pi r^2 \Rightarrow \frac{1}{3}\pi r^2 = (-2t) \times \frac{dt}{dh} \Rightarrow$

$\frac{1}{3}\pi\left(\frac{50\sqrt{3}}{3}\right)^2 = -2t\frac{dt}{dh} \Rightarrow \int\frac{\pi}{3}\left(\frac{50\sqrt{3}}{3}\right)^2 dh = -2t\,dt \Rightarrow \frac{2500}{9}h = -2t + c$.

Then use $h = 50$ when $t = 0$ to find c and finally substitute $h = 0$ to find t.

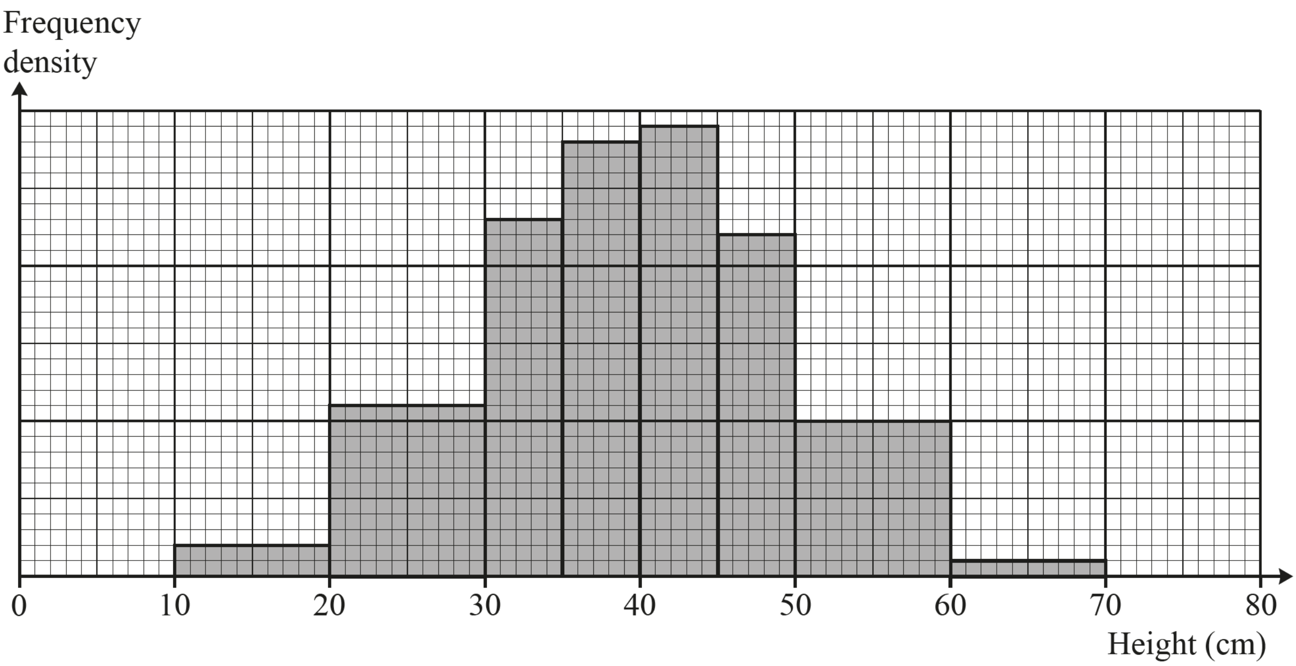
Finally, some candidates used a different, correct method that did not require the use of the chain rule. It did, however, involve some rather awkward substitution of a limit. This method can be found on the published mark scheme, Question 8, Example method 3.

Section B overview

Candidates were generally able to extract information from a statistical diagram or table fairly well. They also coped well with the simple probability Question 13 (a), although less well on the more demanding probability question (13 (b)). Most candidates had clearly been taught how to carry out a hypothesis test, and many did this perfectly.

Question 9 (a)

9 The heights, in centimetres, of a random sample of 150 plants of a certain variety were measured. The results are summarised in the histogram.



One of the 150 plants is chosen at random, and its height, X cm, is noted.

(a) Show that $P(20 < X < 30) = 0.147$, correct to **3** significant figures. **[2]**

In Question 9, some candidates had difficulty dealing with the fact that no scale is given on the Frequency density axis. In fact, so long as area rather than height is used, all the parts can be answered without determining the scale.

Part (a) was answered very well, with candidates using various units with which to measure the areas in the histogram (for example small squares, 5x5 squares, 5×1 rows or cm^2).

Question 9 (b) (i)

Sam suggests that the distribution of X can be well modelled by the distribution N(40, 100).

(b) **(i)** Give a brief justification for the use of the normal distribution in this context. **[1]**

Many candidates gained the mark easily by stating that the histogram is roughly "bell-shaped". The other way to gain the mark was more difficult. Three aspects of the histogram needed to be mentioned (symmetry; highest in the middle; tailing off at each end). Many candidates gave one or two of these, but few gave all three.

Many candidates gave answers that were either inadequate or inappropriate. Some examples were as follows.

"The data is continuous." "The values are discrete." "Data is concentrated around the mean." "The graph is shaped like a normal distribution." "Because nature usually follows a normal curve." "You would expect the heights of plants to be normally distributed." "*n* is large." Even the Central Limit theorem received an occasional mention!

Question 9 (b) (ii)

(ii) Give a brief justification for the choice of the parameter values 40 and 100. **[2]**

Many candidates pointed out that 40 is roughly in the middle, justifying the use of 40 as the mean. For the variance, many candidates used one of the following properties of the normal distribution. Roughly $\frac{2}{3}$ (or 68%) of the data lie within σ of the mean, most of the data lie within 2σ of the mean or all (or almost all) of the data lie within 3σ from the mean. However, some candidates who used the first of these, simply stated that

"Roughly $\frac{2}{3}$ of the data lie within σ of the mean", without actually showing that this was true for the data in the histogram. This was insufficient to gain the mark. A few candidates wrote that $\sigma = 10$ because 10 is the class width.

Some candidates calculated the exact values of the mean and standard deviation. This was unnecessary, but, if correct, gained the marks.

Question 9 (c)

(c) Use Sam's model to find $P(20 < X < 30)$. **[1]**

This was answered correctly by almost all candidates. A few showed working, although this was not necessary.

Question 9 (d)

Nina suggests a different model. She uses the midpoints of the classes to calculate estimates, m and s, for the mean and standard deviation respectively, in centimetres, of the 150 heights. She then uses the distribution $N(m, s^2)$ as her model.

(d) Use Nina's model to find $P(20 < X < 30)$. **[4]**

This part was found difficult because it involved finding the areas of all the blocks in the histogram. Some candidates used the heights, rather than the areas. Having found the frequencies, many candidates showed working for calculating the mean and standard deviation. This was not necessary, and many made errors such as Σf^2 instead of Σfx^2. Other candidates found the frequencies and inputted them into the relevant statistical function on the calculator to give the mean and standard deviation directly. These candidates were usually successful.

If the mean and standard deviation were found correctly, the probability could be found directly from the calculator, although some candidates just found the mean and standard deviation and did not go on to find the probability.

Question 9 (e) (i)

(e) **(i)** Complete the table in the Printed Answer Booklet to show the probabilities obtained from Sam's model and Nina's model. **[2]**

The probabilities for Sam's model were often given correctly. Those for Nina were often incorrect. Many candidates omitted Nina's probabilities and some omitted Sam's also.

Question 9 (e) (ii)

(ii) By considering the different ranges of values of X given in the table, discuss how well the two models fit the original distribution. **[2]**

The wording in the question began "By considering the different ranges of values of X given in the table..." Consequently, candidates needed to make explicit reference to at least two ranges. Many candidates did not do this, but wrote general comments such as "Sam's model is a better fit than Nina's". The published mark scheme shows the responses that were acceptable.

Question 10 (a) (i), (ii) and (iii)

10 The table shows the age structure of usual residents of 18 Local Authorities (LAs) in the North West region of the UK in 2011.

Local Authority	Age 0 to 17	Age 18 to 24	Age 25 to 64	Age 65 and over
A	26.20%	9.06%	51.81%	12.92%
B	23.32%	8.99%	52.32%	15.37%
C	22.24%	8.96%	52.56%	16.23%
D	22.67%	8.10%	53.27%	15.96%
E	20.70%	7.77%	54.77%	16.76%
F	18.14%	6.51%	51.13%	24.21%
G	18.96%	14.20%	48.51%	18.33%
H	19.06%	14.79%	52.12%	14.04%
I	25.15%	9.04%	51.16%	14.65%
J	22.93%	8.81%	52.22%	16.04%
K	21.48%	13.98%	50.82%	13.73%
L	23.98%	9.20%	52.26%	14.56%
M	21.67%	11.19%	52.94%	14.19%
N	17.82%	6.01%	51.93%	24.23%
O	22.83%	7.30%	53.86%	16.01%
P	21.76%	8.28%	54.03%	15.93%
Q	21.42%	8.43%	53.90%	16.25%
R	18.61%	7.33%	49.35%	24.71%

Percentage of residents

(a) Without reference to any other columns, explain how you would use **only** the columns for the age ranges 0 to 17 and 18 to 24 to decide whether an LA might be one of the following.

(i) An LA that includes a university **[1]**

(ii) An LA that attracts young couples to live **[1]**

(iii) An LA that attracts retired people to live **[1]**

Questions 10 (a) (i), (ii) and (iii) left room for a certain amount of interpretation and so were marked generously. The published mark scheme shows the responses that were acceptable. Some candidates did not read the specific instruction "explain how you would use **only** the columns...", and just named a particular Local Authority as their answer to each part.

Question 10 (b) (i)

(b) Using your answers to part **(a)**, identify the following.

(i) Four LAs that might include a university **[1]**

Most answered this question correctly. The four relevant LAs are clearly distinct from the rest. However, a few candidates gave G, H, K and L, presumably intending M rather than L, but having misread the table. Some gave G, H, K and A, but it is not clear why they chose A.

Question 10 (b) (ii)

(ii) Three LAs that might be attractive to retired people **[1]**

Most answered this question correctly also. Again, the three relevant LAs are quite clearly identifiable. However, some candidates gave, for example, F, N and O, or F, N and E although the rationale behind the choice of O or E is difficult to discern. Others gave four LAs, usually F, N, R and O, having, presumably, not read the question carefully enough.

Question 10 (c)

(c) Explain why your answer to part **(b)(ii)**, based only on the columns for the age ranges 0 to 17 and 18 to 24, may not be reliable. **[1]**

Perhaps even more than parts (a) (i), (ii) and (iii), this question left room for much interpretation and so was marked sympathetically. Some candidates' answers were too general, and did not really address the limitation of using only the first two columns to choose the three LAs in part (b) (ii). In some cases answers did not refer to retired people at all. Some examples of inadequate answers were as follows.

"There may be few young couples."

"LAs N and R have a high percentage of young people. Young families could have moved there."

"No information to judge if any LAs are outliers for percentage of 0 – 17 or 18 – 24 year olds."

Examples of acceptable types of answers can be seen in the published mark scheme.

Question 10 (d)

(d) The lower quartile, median and upper quartile of the percentages in the column "Age 65 and over" are 14.56%, 15.99% and 16.76% respectively.

Use this information to comment on your answers to part **(b)(ii)** and part **(c)**. **[2]**

The question is quite specific. A correct answer had to use the information given in this part, and had to refer to both parts (b) (ii) and (c). Most candidates omitted any mention of part (c).

For the first mark, it was not sufficient to say that the relevant LAs' percentages were above the upper quartile. A proper test for outliers was needed, i.e. use of UQ + 1.5 × IQR = 20.06. Knowledge of this test for outliers is required by the syllabus. A few candidates, incorrectly, used median + 1.5 × IQR.

Question 10 (e)

In a magazine article, a councillor plans to describe a typical LA in the North West region. He wants to quote the average percentage of residents aged 65 or over.

(e) The mean of the percentages in the column "Age 65 and over" is 16.90%.

Use this information, and the information given in part **(d)**, to explain whether the median or the mean better represents the data in the column "Age 65 and over". **[2]**

Again, the question is quite specific. Candidates had to use the information given in this part and in part (d), i.e. they had to comment that the mean is greater than the upper quartile. Most candidates gave sensible responses, but did not use the value of the mean given in the question, and so could not score any marks. An example of a reasonable, but inadequate, response was "The median is better because it is less affected by the outliers, LAs F, N and R." or "The median is better because the percentages for LAs F, N and R are above the upper quartile." Even less satisfactory was the standard, rote response "The median is better because it is less affected by outliers."

Question 11

11 In the past the masses of new-born babies in a certain country were normally distributed with mean 3300 g. Last year a publicity campaign was held to encourage pregnant women to improve their diet.

Following this campaign, it is required to test whether the mean mass of new-born babies has increased. A random sample of 200 new-born babies is chosen, and it is found that their mean mass is 3360 g. It is given that the standard deviation of the masses of new-born babies is 450 g.

Carry out the test at the 2.5% significance level. **[7]**

Misconception

Two misunderstandings of the nature of a hypothesis test, such as in Question 11, were shown by some candidates.

The hypotheses are about the proposed mean of the whole population (in this case $\mu = 3300$). But some candidates used the particular value of the sample mean (3360) in their hypotheses (H_0: $\mu = 3360$, H_1: $\mu > 3360$). Then, instead of using $\overline{X} \sim N(3300, \frac{450^2}{200})$ to find $P(\overline{X} > 3360)$, they used $\overline{X} \sim N(3360, \frac{450^2}{200})$ to find $P(\overline{X} < 3300)$, which gives the same answer (0.0297) as the correct method.

Other candidates showed a misunderstanding of the alternative hypothesis, giving hypotheses as follows: H_0: $\mu = 3300$, H_1: $\mu = 3360$.

This question was very well answered by most candidates, more so than similar questions in the past.

The mark scheme for this question (as for all hypothesis test questions) is long and complicated and would reward careful study.

Several candidates gave completely correct solutions, except that they lost just 1 or 2 marks. This was usually caused either by not defining μ in the hypotheses, or by omitting to show explicitly the comparison with 0.025, or by giving a "definite" conclusion such as "The mean mass of new-born babies has not increased." (The conclusion must refer to "evidence".)

A few candidates, although they showed familiarity with the relevant terminology, showed little understanding of the process required for a hypothesis test.

The specification (in paragraph 2.05a) states that the hypotheses for a hypothesis test must be stated in terms of the relevant parameter. In addition to this, the parameter must be defined in context and must be clearly described as the population value. In this question, many candidates did not define the relevant parameter clearly, thus losing a mark. Others gave the hypotheses in words, not in terms of the parameter.

Various methods are possible in this question, including finding a probability or using a critical value of $\overline{X}$ or of z. The latter two methods involve using the formula $\frac{\overline{X}-3300}{450/\sqrt{200}}$, which makes them more time-consuming than just finding $P(\overline{X} > 3360)$.

A common error was to use a bogus "continuity correction", finding $P(\overline{X} > 3360.5$ or $3359.5)$. Candidates who made this error could nevertheless score 6 marks out of 7.

Some candidates found the correct probability (0.0297) and made the correct explicit comparison, 0.0297 > 0.025, but then drew the wrong conclusion, rejecting H_o. Some gave too "definite" a conclusion. Others wrote that there in insufficient evidence that the mean mass of babies has "changed" (rather than "increased").

A few candidates found the correct probability, but gave an incorrect answer of either 0.029 or 0.0296.

Question 12 (a)

12 A firm claims that no more than 2% of their packets of sugar are underweight. A market researcher believes that the actual proportion is greater than 2%. In order to test the firm's claim, the researcher weighs a random sample of 600 packets and carries out a hypothesis test, at the 5% significance level, using the null hypothesis $p = 0.02$.

(a) Given that the researcher's null hypothesis is correct, determine the probability that the researcher will conclude that the firm's claim is incorrect. **[5]**

In this question, probably more than any other, candidates suffered from not reading the question carefully. The question is quite explicit:

"... determine the probability ....". But many candidates thought that the question was asking them to carry out a hypothesis test. Thus they gave hypotheses and a conclusion. In the process of doing this, some candidates carried out some of the processes needed to answer the actual question, and so scored some marks. However, most of these candidates scored only 1 mark (for correctly identifying the distribution B(600, 0.02)).

Those candidates who understood the question correctly generally gave good solutions. Some found $P(X \geq 19) = 0.0359$, then stated that this is less than 0.05, and concluded that the required probability was 0.0359. These scored 4 marks out of 5. In order to gain full marks, candidates had also to show that $P(X \geq 18) = 0.0610$, which is greater than 0.05.

Some candidates appeared to think that the binomial probability function of their calculator gave them, e.g. $P(X < 18)$, whereas it actually gives $P(X \leq 18)$.

A few candidates wrongly assumed that $P(X \geq 18) = 1 - P(X \leq 18)$.

Many candidates correctly identified B(600, 0.02), but then found $np = 12$ and gave $P(X \geq 12) = 0.540$ as their answer.

Some candidates used the normal approximation to the binomial distribution to find the relevant probabilities. These candidates could score a maximum of 4 marks out of 5. But it should be noted that finding probabilities by this method is not in the syllabus.

A few candidates used the normal approximation to the binomial to find roughly which values of X they should consider. Thus: $P(X > a) = 0.05 \Rightarrow X =$ about 17. This was acceptable, but scored only 1 mark unless the candidate then went on to use the binomial distribution to find $P(X \geq 18)$ and $P(X \geq 19)$.

Many candidates carried out correct working and obtained 0.0359, but then took an extra step such as

P(concludes claim incorrect) = $\frac{0.0359}{0.05} = 0.718$.

Question 12 (b)

(b) The researcher finds that 18 out of the 600 packets are underweight. A colleague says

"18 out of 600 is 3%, so there is evidence that the actual proportion of underweight bags is greater than 2%."

Criticise this statement. **[1]**

Most candidates missed the point (which is that P($X \geq$ 18 should be found, rather than just P(X = 18)) giving answers such as "The sample is too small." or "It could have happened by chance and another sample might give a different result." or "3% isn't significantly greater than 2%." or "Sample may not be representative of all bags of sugar."

Question 13 (a)

13 There are 25 students in a class.

- The number of students who study both History and English is 3.
- The number of students who study neither History nor English is 14.
- The number of students who study History but not English is three times the number who study English but not History.

(a)
- Show this information on a Venn diagram.
- Determine the probability that a student selected at random studies English. **[4]**

This question was very well answered by almost all candidates. A few placed History and English the wrong way round in their Venn diagram. A few others gave a correct Venn diagram, but found P(the student studies only English), i.e. $\frac{2}{25}$ instead of $\frac{5}{25}$. Some gave a correct Venn diagram but omitted to find the probability.

Question 13 (b)

Two different students from the class are chosen at random.

(b) Given that exactly one of the two students studies English, determine the probability that exactly one of the two students studies History. **[6]**

A number of different correct approaches to this problem were seen, some of them quite ingenious.

The most common method that was reasonably successful was to use the conditional probability formula: $\frac{\mathrm{P}(\text{exactly one E and exactly one H})}{\mathrm{P}(\text{exactly one E})}$.

The correct working for the denominator is P(Exactly one English) = P(EH') = $\frac{5}{25}\times\frac{20}{24}$ ×2,

and for the numerator is P(exactly one E and exactly one H) = P(HE' and $\mathrm{H}'\mathrm{E}$) + P(EH and $\mathrm{E}'\mathrm{H}'$) = ($\frac{6}{25}\times\frac{2}{24}$ + $\frac{3}{25}\times\frac{14}{24}$) ×2.

Many candidates used this method correctly, except for one or both of the following errors.

They used 25 instead of 24 (i.e. "with replacement").

They omitted the "× 2" (i.e. ignoring the order in which the two students are chosen).

If the method was otherwise correct, these candidates scored 4 marks out of 6.

Some candidates attempted the above method for the numerator, but only included one of the products, either $\frac{6}{25}\times\frac{2}{24}$ or $\frac{3}{25}\times\frac{14}{24}$.

Many candidates made errors involving which regions in the Venn diagram to include. As examples of this, here are some incorrect attempts to calculate P(HE' and $\mathrm{H}'\mathrm{E}$) + P(EH and $\mathrm{E}'\mathrm{H}'$).

($\frac{6}{25}\times\frac{2}{24}$ + $\frac{9}{25}\times\frac{14}{24}$) ×2 (Here the $\frac{9}{25}$ is P(H) where it should be $\frac{3}{25}$ which is P(EH))

$\frac{5}{25}\times\frac{9}{24}$ + $\frac{5}{25}\times\frac{2}{24}$,

$\frac{2}{25}\times\frac{9}{24}$ + $\frac{3}{25}\times\frac{8}{24}$ + $\frac{6}{25}\times\frac{5}{24}$ + $\frac{3}{25}\times\frac{4}{24}$,

$\frac{9}{25}\times\frac{16}{25}$ × 2

Some candidates attempted to use probabilities that applied to only one student. For example $\frac{\mathrm{P(E\ and\ H)}}{\mathrm{P(E)}}$ = $\frac{3/25}{5/25}=\frac{3}{5}$.

This result can, of course, be written down immediately, and some candidates did this as the first step towards another, rather elegant, method that uses conditional probability but not the conditional probability formula:

P(E only / E) × P(H only / not E) + P(E&H / E) × P(Neither / not E) = $\frac{2}{5}\times\frac{6}{20}$ + $\frac{3}{5}\times\frac{14}{20}$ = $\frac{27}{50}$.

Some candidates appeared to be attempting something like this method, but muddled it with the above correct method, showing working such as $\frac{2}{5}\times\frac{6}{24}$ + $\frac{3}{5}\times\frac{14}{24}$.

Supporting you

Post-results services

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.

Keep up-to-date

We send a weekly roundup 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 on the relevant subject page on our website or visit OCR professional development.

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.

It allows you to:

- review and run analysis reports on exam performance
- analyse results at question and/or topic level
- compare your centre with OCR national averages
- identify trends across the centre
- facilitate effective planning and delivery of courses
- identify areas of the curriculum where students excel or struggle
- help pinpoint strengths and weaknesses of students and teaching departments.

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**
- **/ocrexams**
- **/ocrexams**
- **/company/ocr**
- **/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