Qualification Accredited

OCR
Oxford Cambridge and RSA

A LEVEL

Examiners' report

FURTHER MATHEMATICS B (MEI)

H645

For first teaching in 2017

Y432/01 Summer 2024 series

Version 1

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 Y432 series overview

This is the Statistics Minor option component of the MEI Further Mathematics course.

This component was accessible to most candidates who typically attempted to answer all questions.

Questions involving calculations (Questions 1, 2 (b), 2 (c), 3 (d), 4 (b), 4 (d), 5 (b) (i), 5 (b) (ii), 5(d)) were usually done well. These accounted for about one third of the marks in the component.

For the two questions involving hypothesis tests (Questions 3 (e) and 4 (c)) candidates demonstrated a good understanding of the required structure. Hypotheses were stated, critical values were found, and appropriate comparisons made with test statistics. Conclusions were then usually suitably non-assertive and referred to the context of the question. To improve further, candidates should make sure that the hypotheses are accurately expressed and fully in context.

Candidates did not perform as well on 'explain' and 'comment' questions (Questions 3 (b), 3 (c), 4 (a), 4 (e), 4 (f), 5 (a), 5 (c) and 5 (d)). Centres should continue to make sure that sufficient attention is given to these questions when preparing candidates. The conditions required for a probability distribution to be an appropriate model, and the conditions for a test to be valid, need to be clearly understood and expressed in the context of the question.

Candidates needed to realise how the mean and variance of a sum of independent random variables, and of a multiple of a single random variable, could be applied in a practical context, (Question 5 (c), 5 (d)). Candidates needed to identify the questions where pure mathematical methods were required and needed to employ them in unfamiliar contexts (Question 2 (c) and Question 6 (b) (ii)).

Candidates should be encouraged to write neatly. A number of responses to the 'explain' and 'comment' questions were challenging to read.

Useful information regarding sampling, bivariate data and modelling assumptions can be found in the specification for this component.

Candidates who did well on this paper generally:	Candidates who did less well on this paper generally:
• could apply pure mathematical techniques in unfamiliar contexts • could apply expectation algebra in a practical context • had a good understanding of the underpinning concepts of hypothesis testing • could describe and analyse data perceptively with specific references to the data values themselves • showed a good understanding of limitations and assumptions associated with various statistical processes.	• carried out standard probability calculations but not always to the level of accuracy required • carried out basic test procedures reliably, possibly expressing their hypotheses incompletely or inaccurately • found it difficult to identify the key points in 'explain' and 'comment' questions • did not show sufficient working in 'Show that' and 'Determine' questions.

OCR support

A review of the command words used in questions, and the expected level of written response, is outlined in section 2b of the specification. A useful classroom poster, and student version, is available on Teach Cambridge.

Question 1 (a)

1 When a footballer takes a penalty kick the result is that either a goal is scored or a goal is not scored.

It is known that, on average, a certain footballer scores a goal on 85% of penalty kicks. During one practice session, the footballer decides to take penalty kicks until a goal is **not** scored. You may assume that the outcome of each penalty kick that the footballer takes is independent of the outcome of each other penalty kick.

The random variable representing the number of penalty kicks up to and including the first penalty kick that does **not** result in a goal is denoted by X.

(a) State **one** further assumption that is necessary for X to be modelled by a Geometric distribution. **[1]**

This question was answered correctly by most candidates, stating there was a constant probability of scoring or of not scoring.

Some said attempts were independent, even though this assumption was given in the question. Some made the mistake of saying the number of attempts was indefinite, or that X was random.

Question 1 (b)

For the remainder of this question you may assume that this assumption is valid.

(b) Find each of the following.

- $\mathrm{E}(X)$
- $\mathrm{Var}(X)$ **[2]**

This question was answered correctly by most candidates. A few confused the probabilities, using

p = 0.85 rather than *p* = 0.15 and so were only given 1 mark out of 2.

Question 1 (c)

(c) Find the probability that the footballer takes exactly 3 penalty kicks. **[2]**

This question was answered correctly by most candidates. Again, a few confused the probabilities, using

p = 0.85 rather than *p* = 0.15 and so were only given 1 mark out of 2.

Question 1 (d)

(d) Find the probability that the footballer takes at least 5 penalty kicks. **[2]**

This question was answered correctly by most candidates. Again, a few confused the probabilities, using

$p = 0.85$ rather than $p = 0.15$ and so were only given 1 mark out of 2.

Some calculated $P(X>5) = 0.85^5$ not $P(X>4) = 0.85^4$. It is recommended that candidates deal with these integer inequalities using a number line to avoid this kind of error.

Question 2 (a)

2 The sides of a fair 12-sided spinner are labelled 1, 2, ..., 12.

The spinner is spun and X is the random variable denoting the number on the side of the spinner that it lands on.

(a) Suggest a suitable distribution to model X. You should state the value(s) of any parameter(s). **[2]**

This question was answered correctly by most candidates. Almost all used the word 'Uniform'. Most then went on to specify adequately the range of discrete values.

Question 2 (b)

(b) Find each of the following.

- $E(X)$
- $Var(X)$ **[2]**

This question was answered correctly by most candidates.

Question 2 (c)

You are given that E(X) is denoted by μ and Var(X) is denoted by σ^2.

(c) Determine $\mathrm{P}\left(\left|\frac{2(X-\mu)}{\sigma}\right| > 1\right)$. **[3]**

In successful attempts, candidates noted the command word 'Determine' in the question and set out their working clearly, dealing with the modulus inequality correctly and obtaining numerical decimal values before identifying the relevant integers. See Exemplar 1 below.

Another successful method was to set up a table of values for *n* =1 to 12 and calculate sufficient relevant values from the left hand side of the original inequality to identify the correct set of integers.

Many candidates ignored the modulus sign and obtained a one-sided inequality only. Some calculated the correct decimal values but concluded that *X* lay between them, not outside them. Some candidates identified the correct integers to obtain the correct answer but did not show enough evidence of their method.

Exemplar 1

2(c)

$\sigma = \sqrt{11.92} = 3.45$

$\left| \dfrac{2(x-6.5)}{3.45} \right| > 1$

$2(x-6.5) > 3.45$

$|2x-13| > 3.45$

$13-2x > 3.45$ $\qquad 2x-13 > 3.475$

$-2x < -9.55$ $\qquad x > 8.2375$

$x < 4.775$

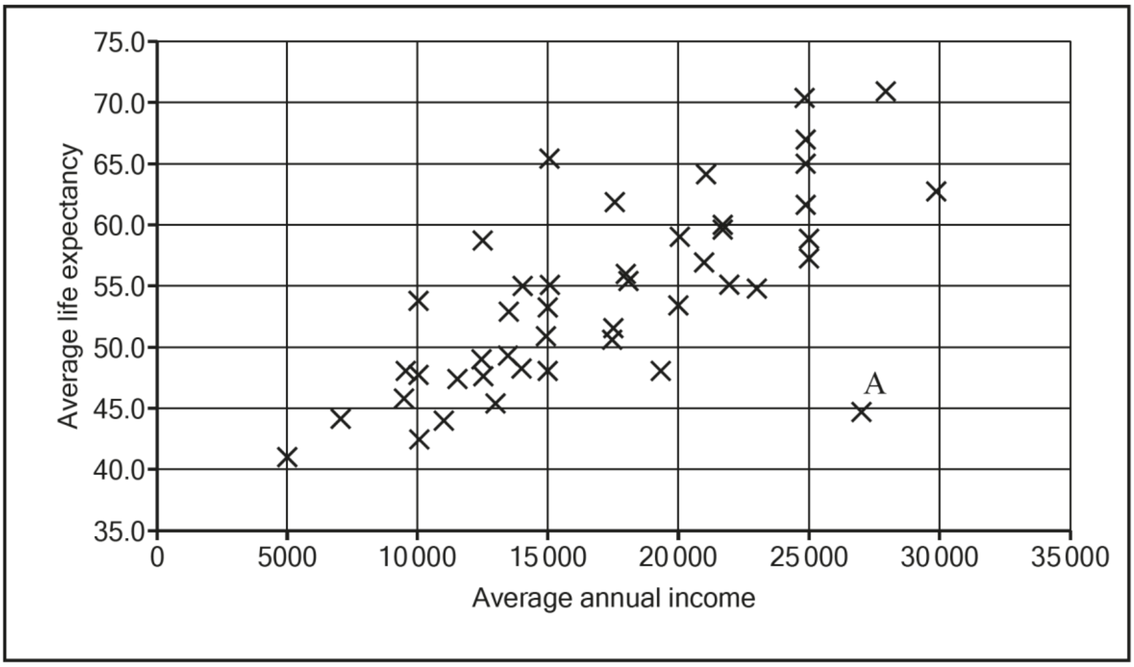
$X = 1, 2, 3, 4, 9, 10, 11, 12$

$\dfrac{8}{12} = \dfrac{2}{3}$

This exemplar shows a clear method of dealing with the modulus inequality (losing the modulus on the second line of working but regaining it on the third), then identifying the correct integers, leading to a fully correct answer as a fraction.

Question 3 (a)

3 The scatter diagram below illustrates data concerning average annual income per person, $\$x$, and average life expectancy, y years, for 45 randomly selected cities.



(a) State whether neither variable, one variable or both variables can be considered to be random in this situation. **[1]**

Most candidates answered correctly.

Question 3 (b)

A student is researching possible positive association between average annual income and average life expectancy. The student decides that the data point labelled A on the scatter diagram is an outlier.

(b) Describe the apparent relationship between average annual income and average life expectancy for this data point relative to the rest of the data. **[1]**

Some candidates recognised what the question was asking and gave a convincing answer, comparing A's values to the rest of the data-points. Many commented inappropriately on the outlier as showing a trend, or as showing a negative correlation, even though it was a single point. See Exemplar 2 below. A few ignored what was asked and simply commented on the rest of the points as showing a positive correlation.

Exemplar 2

3(b)	A ~~shows that~~ suggests that the higher the average annual income, the lower the average life expectancy (negative correlation). ~~B~~ the rest of the data suggests that the higher the average annual income, the higher the average life expectancy (positive correlation)

The response in Exemplar 2 implies that A is part of a general trend, displaying a 'negative correlation', rather than describing its qualities as a single data-point.

Question 3 (c)

The data for point A is removed. The student now wishes to carry out a hypothesis test using the product moment correlation coefficient for the remaining 44 data points to investigate whether there is positive correlation between average annual income and average life expectancy.

(c) Explain why this type of hypothesis test is appropriate in this situation. Justify your answer. **[2]**

Most candidates gave one of the conditions correctly, stating that the data-points, minus the outlier, were elliptically distributed implying a possible bivariate normal distribution. Very few then went on to give the other condition, that the data was random-on-random, despite the clue of part 3(a).

Question 3 (d)

The summary statistics for these 44 data points are as follows.

$\sum x = 751\,120 \quad \sum y = 2397.1 \quad \sum x^2 = 14\,363\,849\,200 \quad \sum y^2 = 133\,014.63 \quad \sum xy = 42\,465\,962$

(d) Determine the value of the product moment correlation coefficient. **[4]**

Almost all candidates showed sufficient working and calculated the correct value. A few incorrectly used $n = 45$ in their calculation losing a mark. A few showed no working at all for which they got no credit as the question clearly said 'Determine'.

Question 3 (e)

(e) Carry out the test at the 1% significance level. **[5]**

Most candidates made a very good attempt at the test. They stated appropriate hypotheses either in terms of rho or in words. In almost all cases they obtained a correct critical value followed by a correct decision about H_0. Then most gave a suitable non-assertive conclusion. This seems to indicate an improvement in candidate preparation over the last few years.

In their hypotheses, a few lost marks for omitting the word 'population' or for lack of context. There are still a few whose conclusion was too assertive.

Some used the word 'association' instead of 'correlation' in both their written hypotheses and their conclusion. It needs to be made clear that a test involving the correlation coefficient must be a test for correlation, not anything else.

Question 4 (a)

4 A genetics researcher is investigating whether there is any association between natural hair colour and natural eye colour. A random sample of 800 adults is selected. Each adult can categorise their natural hair colour as blonde, brown, black or red and their natural eye colour as brown, blue or green.

(a) Explain the benefit of using a **random** sample in this investigation. **[1]**

Most candidates realised that a random sample is not only unbiased but also has a purpose. Many said that it could be used for 'proper inferences' which was acceptable. Some lost the mark for only saying that it would be 'representative' or would 'reduce bias'. Others lost it for saying 'inferences can be made' without any kind of validating word or expression about accuracy.

Question 4 (b)

The data collected from the sample are summarised in **Table 4.1**.

Table 4.1

Observed frequency		Hair Colour				
		Blonde	Brown	Black	Red	Total
Eye Colour	Brown	47	153	196	36	432
	Blue	61	78	115	26	280
	Green	19	22	31	16	88
	Total	127	253	342	78	800

The researcher decides to carry out a chi-squared test.

(b) Determine the expected frequencies for each eye colour in the blonde hair category. **[2]**

Almost all candidates showed appropriate working for at least one calculation and then found all the correct answers to 2 decimal places. A few showed no working. In this case the marks were earned only if all three values were exactly correct. Given that they were asked to 'determine' the values, this was generous and may not be repeated in future papers.

Question 4 (c)

You are given that the test statistic is 28.62 to 2 decimal places.

(c) Carry out the chi-squared test at the 10% significance level. **[5]**

Most candidates made a good attempt at the test. They stated appropriate hypotheses in words. In almost all cases they then calculated the correct number of degrees of freedom followed by the correct critical value. They made the comparison with the given value and reached a correct decision about H_0. Then most gave a suitable non-assertive conclusion. Again, this seems to indicate an improvement in candidate preparation over the last few years.

In their hypotheses, a few lost marks for lack of context. There are still a few whose conclusion was too assertive. A few used the word 'correlation' instead of 'association' in the conclusion. This and a similar error in 3(e) indicate that some candidates are confusing the two.

Question 4 (d)

Table 4.2 shows the chi-squared contributions for some of the categories. The contributions for the categories relating to green eye colour have been deliberately omitted.

Table 4.2

Chi-squared contributions		Hair Colour			
		Blonde	Brown	Black	Red
Eye Colour	Brown	6.791	1.964	0.694	0.889
	Blue	6.162	1.257	0.185	0.062
	Green				

(d) Calculate the chi-squared contribution for the green eye and blonde hair category. **[1]**

Almost all candidates made the correct calculation. A few lost the mark for writing their answer to 4 decimal places when clearly the other values in Table 4.2 are given to 3 decimal places. Candidates need to remember that the *appropriate* level of accuracy is required.

Question 4 (e)

(e) With reference to the values in **Table 4.2**, discuss what the data suggest about brown eye colour and blue eye colour for people with blonde hair. **[2]**

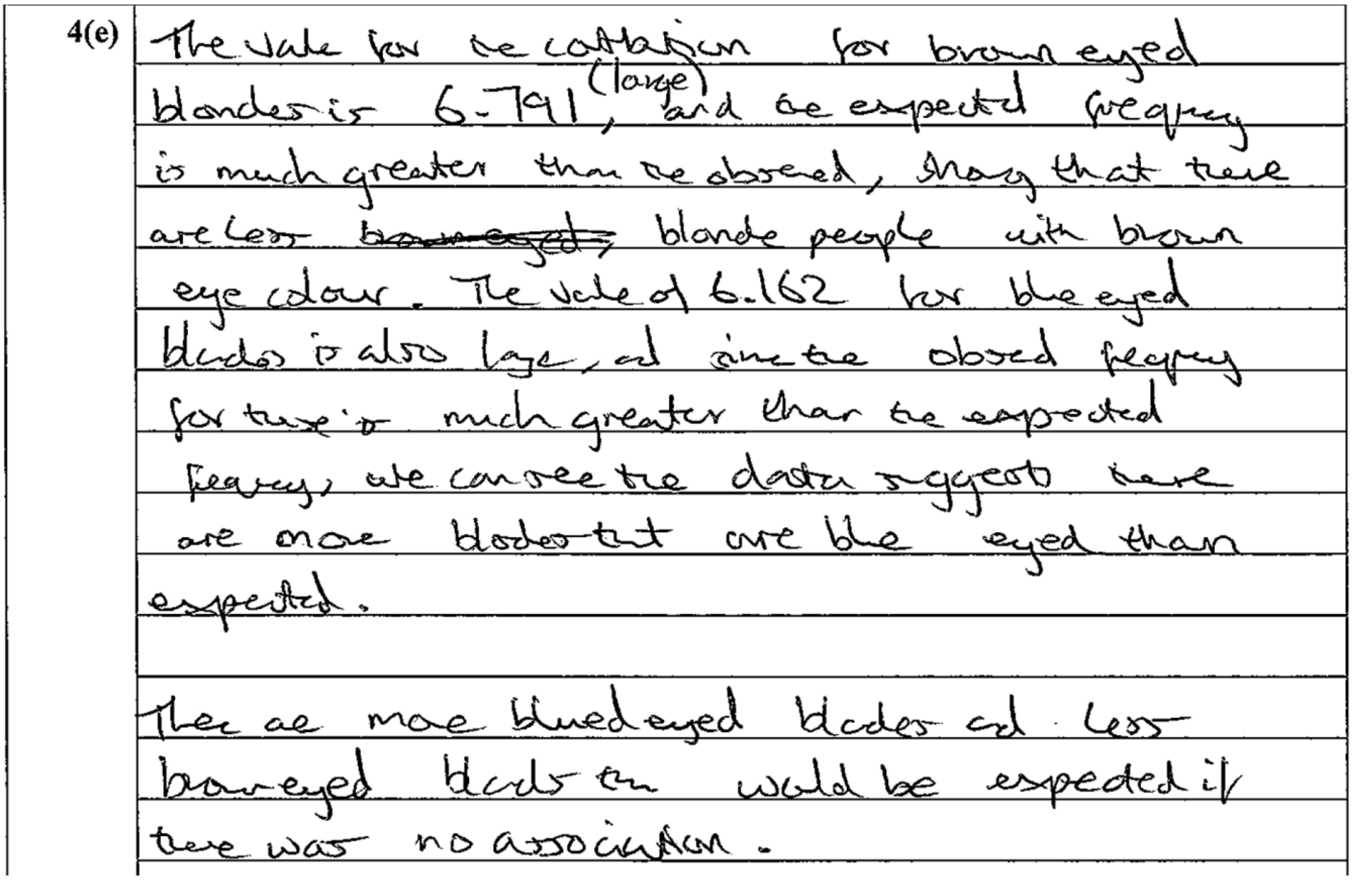
Many candidates referred to the chi-squared contributions in Table 4.2 and commented that their magnitude suggested a significant difference between observed and expected frequencies for brown and blue eyes. This was sufficient for the first mark, even when the values of the contributions were not quoted. Very few candidates then went on to quote and compare observed and expected frequencies for the two eye colours, which was required for the second mark.

Some candidates lost the first mark for stating that there were more brown eyes than expected. A few did not refer to Table 4.2 at all but only compared frequencies thus receiving no credit.

Candidates sometimes wrote at length, without supporting their statements with quoted values and to little effect.

Exemplar 3 below is a typical response to this part. It gains the first mark for mentioning and quoting the contributions and saying they imply a significant difference between observed and expected frequencies, and the inequalities mentioned are the right way round, but those frequencies are not quoted to provide evidence. The candidate has written at some length but only received 1 mark out of 2.

Exemplar 3


4(e) The value for the contribution for brown eyed blondes is 6.791 (large), and the expected frequency is much greater than the observed, showing that there are less ~~brown eyed~~ blonde people with brown eye colour. The value of 6.162 for blue eyed blondes is also large, and since the observed frequency for these is much greater than the expected frequency, we can see the data suggests there are more blondes that are blue eyed than expected.

There are more blue eyed blondes and less brown eyed blondes than would be expected if there was no association.

Assessment for learning

Comments analysing or describing data need to be succinct and structured. Every point that is made should be supported by quoting values or facts drawn directly from the data.

Question 4 (f)

(f) A different researcher, carrying out the same investigation, independently takes a different random sample of size 800 and performs the same hypothesis test, but at the 1% significance level, reaching the same conclusion as the original test.

By comparing only the significance level of the two tests, specify which test, the one at the 10% significance level or the one at the 1% significance level, provides stronger evidence for the conclusion. Justify your answer. **[1]**

Many candidates answered this successfully, referring correctly to the percentages as the chance that H_0 is rejected when true, or as the chance of a false positive, and so comparing them.

Many candidates showed a lack of understanding of the significance level, simply saying that the lower percentage meant a stronger test or was harder to achieve. Some said that the percentage was the chance of rejecting the wrong hypothesis but did not specify which hypothesis.

Question 5 (a)

5 Over a long period of time, it is found that the mean number of mistakes made by a certain player when playing a particular piece of music is 5.

The number of mistakes that the player makes when playing the piece is denoted by the random variable Y.

(a) State **two** assumptions necessary for Y to be modelled by a Poisson distribution. **[2]**

Many candidates answered successfully with some context for both marks. Most got at least one assumption correct, often quoting the acceptable phrase 'mistakes occur at a constant average rate'. There is still a problem with phrasing the condition of independence. Candidates need to say, 'mistakes occur independently' or 'the event a mistake happens is independent of another mistake happening', and should avoid the error of saying the 'number of mistakes is independent'.

Question 5 (b) (i)

For the remainder of this question you may assume that Y **can** be modelled by a Poisson distribution.

(b) (i) Find the probability that the player makes exactly 3 mistakes when playing the piece. **[1]**

This question was answered correctly by most candidates.

Question 5 (b) (ii)

(ii) Find the probability that the player makes fewer than 3 mistakes when playing the piece. **[1]**

This question was answered correctly by most candidates.

Question 5 (b) (iii)

(iii) Find the probability that the player makes fewer than 6 mistakes in total when playing the piece twice, assuming that the performances are independent. **[2]**

This question was answered correctly by most candidates.

Some candidates realised that the combined distribution was Po(10) but found the wrong probability for 1 mark out of 2. A few multiplied two probabilities obtained from Po(5) for no marks.

Question 5 (c)

In a recording studio, the player plays the piece once in the morning and once in the afternoon each day for one week (7 days). It can be assumed that all the performances are independent of each other.

The performances are recorded onto two CDs, one for each of two critics, A and B, to review. The critics are interested in the total number of mistakes made by the player **per day**. Unfortunately, there is a recording error in one of the CDs; on this CD, every piece that is supposed to be an afternoon recording is in fact just a repeat of that morning's recording.

The random variables M_1 and M_2 represent the total number of mistakes per day for the correctly recorded CD and for the wrongly recorded CD respectively.

(c) By considering the values of $\mathrm{E}(M_1)$ and $\mathrm{E}(M_2)$ explain why it is not possible to use the mean number of mistakes per day on the CDs to determine which critic received the wrongly recorded CD. **[2]**

Very few candidates carried out the required expectation algebra to show that the expected value of $Y_{AM}+Y_{PM}$ was the same as that of $2Y_{AM}$. Many wrote at some length arguing that it was obvious that both values were 10 which gained no credit.

Question 5 (d)

Each critic counts the total number of mistakes made **per day**, for each of the 7 days of recordings on their CD. Summary data for this is given below.

Critic A: $n = 7, \quad \sum x_A = 70, \quad \sum x_A^2 = 812$

Critic B: $n = 7, \quad \sum x_B = 72, \quad \sum x_B^2 = 800$

(d) By considering the values of $\mathrm{Var}(M_1)$ and $\mathrm{Var}(M_2)$ determine which critic is likely to have received the wrongly recorded CD. **[4]**

Very few candidates used expectation algebra to show that $\mathrm{Var}(Y_{AM}+Y_{PM})$ was 10 and $\mathrm{Var}(2Y_{AM})$ was 20. Most started by trying to calculate the variance of the given data values, either σ_n^2 or σ_{n-1}^2. Many were given the M1 mark for a correct attempt at this, but a surprising number seemed to have forgotten how to calculate a variance from summary data. Then many lost the last mark for being unable to identify the right person with a correct reason. A lot of students thought that the wrong disc with a repeated value would have a lower variance not a higher one.

Question 6 (a)

6 The probability distribution of a discrete random variable, X, is shown in the table below.

x	0	1	2
$\mathrm{P}(X = x)$	$1 - a - b$	a	b

(a) Find $\mathrm{E}(X)$ in terms of a and b. **[1]**

Almost all candidates answered correctly.

Question 6 (b) (i)

(b) **(i)** In the case where $\mathrm{E}(X) = a + 0.4$, find an expression for $\mathrm{Var}(X)$ in terms of a. **[3]**

Almost all candidates correctly found the value of *b*, then most went on to find the correct quadratic in *a*.

Question 6 (b) (ii)

(ii) In this case, show that the greatest possible value of $\mathrm{Var}(X)$ is 0.65. You must state the associated value of a. **[2]**

Many candidates used a valid method here, many differentiating and solving equal to zero, some completing the square. Some candidates did not know how to proceed and simply equated their quadratic to 0.65 to solve for *a*. This received no credit as they had not shown that 0.65 was the maximum value of Var(*X*).

Question 6 (c) (i)

(c) You are now given instead that $\mathrm{E}(X)$ is not known.

(i) State the least possible value of $\mathrm{Var}(X)$. **[1]**

Most candidates realised with little working that the minimum value was 0.

A few unsuccessfully attempted an algebraic approach, obtaining complicated expressions in terms of *a* and *b*. A noticeable number of candidates did not attempt an answer.

Question 6 (c) (ii)

(ii) Give all possible pairs of values of a and b which give the least possible value of Var(X) stated in part **(c)(i)**. **[2]**

Many candidates realised that if X takes a single value with a probability of 1 then its variance is 0. They then found at least one correct pair of values of a, b to achieve this. Some candidates were exhaustive in their approach and successfully found all three pairs of values. Again, this was a question part that some candidates did not attempt.

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

We've made it easier for Exams Officers to download copies of your candidates' completed papers or 'scripts'. 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 professional development 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 a free test-building platform, providing unlimited users exclusively for staff at OCR centres with an Interchange account.

Choose from a large bank of questions to build personalised tests and custom mark schemes, with the option to add custom cover pages to simulate real examinations. You can also edit and download complete past papers.

Find out more.

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 (examined units only).

Find out more.

You will need an Interchange account to access our digital products. 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.

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

OCR acknowledges the use of the following content: N/A