

**A LEVEL**

**Examiners' report**

**FURTHER  
MATHEMATICS A**

**H245**

For first teaching in 2017

**Y542/01 Summer 2023 series**

# Contents

Introduction .....3

Paper Y542/01 series overview .....4

    Question 1 (a) .....5

    Question 1 (b) (i) .....6

    Question 1 (b) (ii) .....6

    Question 1 (c) .....7

    Question 1 (d) .....7

    Question 2 (a) .....8

    Question 2 (b) .....8

    Question 2 (c) .....9

    Question 2 (d) .....9

    Question 3 (a) .....10

    Question 3 (b) .....11

    Question 4 (a) .....11

    Question 4 (b) .....12

    Question 4 (c) .....12

    Question 4 (d) .....12

    Question 5 (a) .....13

    Question 5 (b) .....13

    Question 6 .....14

    Question 7 (a) .....16

    Question 7 (b) .....16

    Question 8 (a) .....17

    Question 8 (b) .....17

    Question 8 (c) .....17

    Question 8 (d) .....18

    Question 8 (e) .....18

    Question 8 (f) .....18

    Question 8 (g) .....19

## 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.

### Would you prefer a Word version?

Did you know that you can save this PDF as a Word file using Acrobat Professional?

Simply click on **File > Export to** and select **Microsoft Word**

(If you have opened this PDF in your browser you will need to save it first. Simply right click anywhere on the page and select **Save as . . .** to save the PDF. Then open the PDF in Acrobat Professional.)

If you do not have access to Acrobat Professional there are a number of **free** applications available that will also convert PDF to Word (search for PDF to Word converter).

## Paper Y542/01 series overview

Many candidates had prepared thoroughly for this paper, with only a few indications (such as confusion between two different methods, or perhaps getting hypotheses the wrong way round) that preparation had been in any way restricted. The paper allowed most candidates to score well on at least some questions, while there were undoubtedly challenging aspects to other questions. A substantial number of questions requiring verbal responses shone spotlights on common misunderstandings. Some very clear explanations were seen, but there were also numerous vague or muddled comments. The overall impression was that many candidates can carry out calculations correctly but have only a limited understanding of the ideas and concepts, and that they often simply write down words they have seen in textbooks or on previous mark schemes without having a clear idea of what they mean.

Some of the underlying concepts in this specification are quite subtle. Perhaps candidates would benefit from spending more time discussing examples, in a way that is not often dealt with in textbooks or other readily available sources.

Candidates who did well on this paper generally:	Candidates who did less well on this paper generally:
<ul style="list-style-type: none"> <li>• had a proper understanding of modelling assumptions</li> <li>• avoided confusing the use of the same letter in different statistical contexts</li> <li>• in hypothesis tests, consistently got the hypotheses, and also the conclusions, the right way round</li> <li>• understood that statistical hypotheses referred to the underlying populations</li> <li>• consistently appreciated the differences between sample mean/variance and expected mean/variance; they knew that sample values needed to relate only approximately to the corresponding theoretical equations.</li> </ul>	<ul style="list-style-type: none"> <li>• had learnt modelling assumptions by rote but had not properly grasped their meaning</li> <li>• confused the use of the same letter in different contexts</li> <li>• in hypothesis tests, often got hypotheses and/or conclusions the wrong way round</li> <li>• stated hypotheses in terms of samples or data rather than the underlying populations</li> <li>• confused the relationship between sample mean/variance and expected mean/variance, or assumed that equations for expected values applied exactly to the corresponding sample values.</li> </ul>

## Question 1 (a)

- 1 A certain section of a library contains several thousand books. A lecturer is looking for a book that refers to a particular topic. The lecturer believes that one-twentieth of the books in that section of the library contain a reference to that topic. However, the lecturer does not know which books they might be, so the lecturer looks in each book in turn for a reference to the topic. The first book the lecturer finds that refers to the topic is the  $X$ th book in which the lecturer looks.
- (a) A student says, "There is a maximum value of  $X$  as there is only a finite number of books. So a geometric distribution cannot be a good model for  $X$ ."

Explain whether you agree with the student.

[1]

Successful responses said that, because the number of books is large, the difference between sampling with or without replacement is negligible. The issue is whether a geometric distribution is a *good enough* model, not whether it is perfect. The situation is exactly the same with a binomial distribution, which provides a perfectly adequate model for selection without replacement provided the population (not the sample!) is big enough. Commonly candidates missed the point and gave responses relating only to learnt phrases.

## Assessment for learning



With both the binomial and the geometric distributions, the difference between sampling with and without replacement is often negligible, especially with a large population.

Students might find it helpful to consider numerical examples. Suppose the relevant section of the library has  $n$  books. If  $n = 5000$ , a spreadsheet calculation shows that the values of  $P(X = x)$  are the same to within  $\pm 0.0001$ , whether sampling is with replacement or not, for all values of  $x$ . But if  $n = 100$  the two answers are not so close. For example,  $P(X = 10) = 0.0315\dots$  from the geometric distribution, but  $95/100 \times 94/99 \times 93/98 \times \dots \times 87/92 \times 5/91$  gives  $0.0339\dots$

## Exemplar 1

1(a)	<p>I disagree I agree as the finite number of books means a varying probability of success so the geometric distribution isn't appropriate</p>
------	------------------------------------------------------------------------------------------------------------------------------------------------

This was the most common wrong answer. It was necessary to appreciate that, because the number of available books was large, it made no appreciable difference to use 0.95 throughout.

## Question 1 (b) (i)

- (b) (i) State **one** modelling assumption (not involving the total number of books) needed for  $X$  to be modelled by a geometric distribution in this context. [1]

Successful responses saw candidates give answers such as 'whether or not one book chosen contains a reference to the topic does not affect whether the next book chosen does so', or 'any book in the section is equally likely to contain a reference to the topic'.

Most candidates could give an appropriate assumption, but often stated it badly. Some gave not a modelling *assumption* but a modelling *condition*, such as 'there must be an infinite number of possible trials', or 'books must either refer to the topic or not'. Conditions such as these are rarely of interest, as it is usually obvious whether or not they apply, and in any case they merely give a first indication of which distribution is likely to be appropriate, not whether that distribution is a good model. A response such as 'Whether a book refers to a topic must be independent of whether it does not refer to the topic' shows the sort of confusion that can arise when a candidate attempts to use learnt phrases without any understanding of what they might mean.

## Question 1 (b) (ii)

- (ii) Suggest a reason why this assumption may **not** be valid in this context. [1]

The most obvious reason, chosen by many candidates, was that in any section of a library books are not arranged randomly but grouped by topic, so that it is unlikely that the books are independent of one another. Alternatively, a correct application of the 'constant probability' assumption would be that books on some shelves might be more likely to contain a reference than those on others.

Many said that sampling without replacement meant that the probabilities change, but the whole point of part (a) was to suggest that, with a large number of books, such a change is negligible. For instance, starting with 5000 books and using changing probabilities, the answer to part (c) is still 45.

**Question 1 (c)**

Assume now that  $X$  can be well modelled by the distribution  $\text{Geo}(0.05)$ .

- (c) The probability that the lecturer needs to look in no more than  $n$  books is greater than 0.9.

Find the smallest possible value of  $n$ .

[3]

Many candidates answered this correctly, either by solving  $0.95^n < 0.1$  or by using trial and improvement, while some had 'invGeo' facility on their calculator.

Those who solved  $0.95^n < 0.1$  by taking logarithms to base 0.95 needed to remember that  $\log_{0.95}$  is a *decreasing* function and so the inequality is reversed:  $n > \log_{0.95} 0.1$ .

Some candidates tried to use  $0.95^{n-1}$  or  $0.05 \times 0.95^{n-1}$ . A tiny number of candidates used  $N(20, 380)$ , which happens to give the correct answer in this case but which is not a valid method as geometric distributions cannot be approximated as a whole by normal distributions (their graphs are never symmetric). For example, using  $N(20, 380)$  would give  $P(X \leq 0)$  to be 0.1586, whereas  $X \leq 0$  is impossible for a geometric distribution.

**Question 1 (d)**

- (d) The lecturer needs to find four different books that refer to the topic.

Find the probability that the lecturer wants to look in exactly 40 books.

[2]

Although this is a common enough question in the context of a geometric distribution, it had not been asked on the new specification, which perhaps explains why few correct answers were seen. The method is to find the probability that there are 3 successes in the first 39 trials, and then that the 40<sup>th</sup> trial is a success.

## Question 2 (a)

- 2 The director of a concert hall wishes to investigate if the price of the most expensive concert tickets affects attendance. The director collects data about the price, £ $P$ , of the most expensive tickets and the number of people in the audience,  $H$  hundred (rounded to the nearest hundred), for 20 concerts. For each price there are several different concerts. The results are shown in the table.

$P$ (£)	75	65	55	45	35
$H$ (hundred)	27	27	27	26	15
	27	27	20	21	12
		22	18	16	9
		19	18	13	
		12	16	9	

$$n = 20 \quad \sum p = 1080 \quad \sum h = 381 \quad \sum p^2 = 61300 \quad \sum h^2 = 8011 \quad \sum ph = 21535$$

- (a) Calculate the equation of the regression line of  $h$  on  $p$ . [2]

This was generally well done, although there was some premature approximation (leading to  $a = 1.66$  instead of 1.64), and some candidates gave the answer in terms of  $x$  and  $y$  instead of  $p$  and  $h$ .

## Question 2 (b)

- (b) State what change, if any, there would be to your answer to part (a) if  $H$  had been measured in thousands (to 1 decimal place) rather than in hundreds. [1]

Answers were about equally divided between the correct  $0.0332p + 0.164$  (or 'the coefficients are both divided by 10'), the insufficient 'the coefficients are smaller' and 'no change as it is a linear transformation' (which of course is true of the PMCC but not of the regression equation).



## Question 2 (c)

For a special charity concert, the most expensive tickets cost £50.

- (c) Use your answer to part (b) to estimate the expected size of the audience for this concert. Give your answer correct to 1 decimal place. [1]

Often correct, but many candidates gave answers such as 17.7, instead of 17.7 *hundred*. The question asks for the size of the audience, not for the value of  $h$ .

## Question 2 (d)

- (d) Comment on the reliability of your answer to part (c). You should refer to
- the value of the product-moment correlation coefficient for the data, which is 0.642
  - the value of £50
  - any **one** other relevant factor that should be taken into account. [4]

Many candidates could make two or three valid points here, although some said that it was unreliable as 50 was not one of the given values of  $P$  and others said that it might depend on the lower prices as well; both of these comments overlook the whole point of doing a regression calculation on this data, which is to average the variability out into an overall trend.

In order to score full marks, candidates needed to write a single summative conclusion, such as 'Overall it is not very reliable', and not just comment on the effect of each comment separately. Any response in the range 'not reliable' to 'fairly reliable' inclusive was allowed; candidates were not being tested on whether the examiner agreed with their judgement but on whether they were able to coordinate the different issues.

It should be added that comparison of 0.642 with critical values is not relevant here. Such a comparison would give evidence of whether there *is* any correlation (i.e.,  $\rho > 0$ ), but what we need to know is *how good* the correlation is (i.e., whether  $\rho$  is close to 1).

Exemplar 2

2(d)	<p>A price index of 0.642 suggests the data does approximately fit a linear relationship, modelled by the regression line, suggesting it is reliable.</p> <ul style="list-style-type: none"> <li>• £50 is also within the range of the data meaning we're extrapolating which encourages <del>reliability</del> reliability</li> <li>• The fact that it is a charity concert, however, make more people likely to attend, <del>than for other</del> <sup>than for other</sup> concerts meaning the data may have been biased, reducing reliability</li> </ul>
------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

This is a typical example of a candidate who has made three valid points, but has commented only on each point separately. There needs to be an overall conclusion as to the reliability.

Question 3 (a)

3 The discrete random variable  $W$  has the distribution  $U(11)$ . The independent discrete random variable  $V$  has the distribution  $U(5)$ .

(a) It is given that, for constants  $m$  and  $n$ , with  $m > 0$ ,  $E(mW + nV) = 0$  and  $\text{Var}(mW + nV) = 1$ .

Determine the exact values of  $m$  and  $n$ .

[5]

This was very well done. Correct answers were widely seen.

### Question 3 (b)

The random variable  $T$  is the mean of three independent observations of  $W$ .

- (b) Explain whether the Central Limit Theorem can be used to say that the distribution of  $T$  is approximately normal. [1]

The point here is that three independent observations is insufficient for the CLT to apply. Candidates who said ' $n$  is not greater than 25' had to say what  $n$  was; the need of this was clear from those candidates who took  $n$  to be 11 (from the distribution  $U(11)$ ), or even 33. Candidates need to internalise the meaning of letters such as  $n$  in the context of either the CLT or uniform distributions, and not just to treat  $n$  as a meaningless symbol.

Some added that we are not told that the parent population is normal. That is the answer to a different question: 'is the CLT *needed*?'. It is true that the mean of any number of samples taken from a normal population is itself distributed normally, but that is a different theorem. The CLT is needed only when the parent population is not normal.

### Question 4 (a)

- 4 Two magazines give numerical ratings to hi-fi systems. Li wishes to test whether there is agreement between the opinions of the magazines. Li chooses a random sample of 5 hi-fi systems and looks up the ratings given by the two magazines. The results are shown in the table.

System	A	B	C	D	E
Magazine I	68	75	77	83	92
Magazine II	30	25	40	35	45

- (a) Give a reason why Li might choose to use a test based on Spearman's rank correlation coefficient rather than on Pearson's product-moment correlation coefficient. [1]

Many candidates gave sensible comments such as 'the agreement between the ratings may not be linear' or 'the ratings are subjective'. Another correct response was 'the data might not have a bivariate normal distribution' (but the word 'bivariate' must be included here). However, it is not enough to say that the ratings might be on different scales, as mere scaling of the marks does not change the value of Pearson's coefficient.

In this sort of question it is usually helpful to distinguish between ratings and rankings.

**Question 4 (b)**

- (b) Calculate the value of Spearman's rank correlation coefficient for the data. [4]

This was very often correct, although some candidates attempted to use raw scores and others ranked all 10 ratings from 1 to 10, perhaps in confusion with a Wilcoxon test.

**Question 4 (c)**

- (c) Use your answer to part (b) to carry out a hypothesis test at the 5% significance level. [4]

'Agreement' needs to indicate positive correlation, and this should be stated in the hypotheses. It is therefore a one-tailed test, but quite a lot of candidates used a two-tailed test with a critical value of 1.0.

The correct conclusion is that there is insufficient evidence of agreement between the opinions. It is wrong to say that there is evidence of no agreement. It cannot be stressed often enough that a hypotheses test never produces evidence that the null hypothesis is correct.

**Question 4 (d)**

- (d) The value of Spearman's rank correlation coefficient between the ratings given by magazine I and by a third magazine, magazine III, has the same numerical value as the answer to part (b) but with the sign changed.

In the Printed Answer Booklet, complete the table showing the rankings given by magazine III. [1]

The rank order of the results for magazine II has to be reversed (so that each new rank is 6 minus the old rank). Relatively few correct answers were seen.

### Question 5 (a)

- 5 An historian has reason to believe that the average age at which men got married in the seventeenth century was higher in urban areas compared to rural areas. The historian collected data from a random sample of 8 men in an urban area and a random sample of 6 men in a rural area, all of whom were married in the seventeenth century. The results were as follows, given in the form years/months.

Urban: 18/3    18/5    19/9    20/7    25/6    34/6    41/8    46/3  
 Rural: 18/0    18/1    18/4    19/11    22/2    28/11

- (a) Use an appropriate non-parametric method to test at the 5% significance level whether the average age at marriage of men is higher in urban areas than in rural areas. [7]

This was generally well done and many candidates found it a good source of marks. The hypotheses need to be stated in terms of the (population) *median* ages of marriage; many lost a mark by referring only to 'average' or 'mean'. (A handful used the correct alternative that the distributions have to be identical apart from their average, and in that case any average can be used, but this applies only to statements involving the whole distributions being identical.)

Most candidates remembered to consider not just  $R_m = 34$  but also  $m(m + n + 1) - R_m = 56$ . However, some considered 71, which is wrong; that is the sum of the rankings of the larger group ( $R_n$ ). The alternatives for  $R_m$  are needed to allow for the data to be ranked either top-down or bottom-up, but must always use the smaller group.

The calculations were mostly correct, but the final mark was often lost for the incorrect conclusion 'there is evidence that the average ages are the same'. The conclusion for this test should be 'there is insufficient evidence that the average age of marriage is greater in urban areas'.

### Question 5 (b)

- (b) When checking the data, the historian found that the age of one of the men, Mr X, which had been recorded as 28/11, had been wrongly recorded. When corrected, the result of the test in part (a) was unchanged.

Determine the youngest age that Mr X could have been, given that it was not the same, in years and months, as that of any of the other men in the sample. [3]

Quite a lot of right answers were seen to this question, often by trial and improvement. Some candidates correctly calculated that the old 11<sup>th</sup>-ranked value could become the 8<sup>th</sup> and gave 20/0 as their answer, but in fact it could become the 7<sup>th</sup>. Some candidates simply gave the lowest age that would not change the rankings at all, namely 25/7.

## Question 6

- 6 The continuous random variable  $X$  has a uniform distribution on the interval  $[-\pi, \pi]$ .

The random variable  $Y$  is defined by  $Y = \sin X$ .

Determine the cumulative distribution function of  $Y$ .

[7]

This question illustrated the dangers of proceeding entirely by rote. A common error is seen below. The issue is that the function  $\sin X$  is not 1-1 in the domain  $[-\pi, \pi]$ , so that  $P(\sin X < y)$  is not the same as  $P(X < \sin^{-1}y)$ . For example, the solution to the inequality  $\sin X < 0.5$  is not just  $X < \pi/6$ ; in the given domain it is  $\{X < \pi/6\} \cup \{X > 5\pi/6\}$ . In general it is necessary to consider also  $X > \pi - \sin^{-1}y$ . One or two candidates also realised that the answer was unchanged if the domain was  $[-\pi/2, \pi/2]$  or adjusted their final answers to make sure that  $F(-\pi) = 0$  and  $F(\pi) = 1$ . Those who drew diagrams were more likely to succeed.

The final CDF should be given in its full form, with the ranges where it is equal to 0 or to 1 as well as the substantive part. Many gave the range for the substantive part as  $y = 0$  only, presumably from the values of  $\sin X$  at the end-points of the range, although obviously this cannot be correct.

Exemplar 3

6  $f(x) = \begin{cases} \frac{1}{2\pi} & -\pi \leq x \leq \pi \\ 0 & \text{o/w} \end{cases}$

~~$g(x) = \begin{cases} \frac{\sin^{-1} x}{2\pi} \end{cases}$~~

$P(x) = \begin{cases} \frac{x}{2\pi} + \frac{1}{2}, & -\pi \leq x \leq \pi \\ 0 & x < -\pi \\ 1 & x > \pi \end{cases} = P(X \leq x)$

$G(x) = P(X \leq x)$   
 $= P(\sin X \leq x)$   
 $= P(X \leq \sin^{-1} x) = F(\sin^{-1} x) = \begin{cases} \frac{\sin^{-1} x}{2\pi} + \frac{1}{2} & -\pi \leq \sin^{-1} x \leq \pi \\ 0 & \sin^{-1} x < -\pi \\ 1 & \sin^{-1} x > \pi \end{cases}$

$F(\sin^{-1} x) = \begin{cases} \frac{\sin^{-1} x}{2\pi} + \frac{1}{2}, & -1 \leq x \leq 1 \\ 0, & x < -1 \\ 1, & x > 1 \end{cases}$   
 $= G(x)$   
 $= \text{cdf of } Y$

This candidate has followed the standard procedure correctly, except that the line after  $P(\sin X \leq x)$  ought to read  
 $= P(X \leq \sin^{-1} x) + P(X \geq \pi - \sin^{-1} x)$   
 which then gives  $\frac{2 \sin^{-1} x}{2\pi} + \frac{1}{2}$  (note the extra 2 in the numerator).

## Question 7 (a)

- 7 A club secretary collects data about the time,  $T$  minutes, needed to process the details of a new member. The mean of  $T$  is denoted by  $\mu$ . The variance of  $T$  is denoted by  $\sigma^2$ . The results of a random sample of 40 observations of  $T$  are summarised as follows.

$$n = 40 \quad \sum t = 396.0 \quad \sum t^2 = 4271.40$$

- (a) Determine a 99% confidence interval for  $\mu$ .

[7]

This was largely well done and a good source of marks for many candidates. Most remembered to use  $\sigma/\sqrt{40}$  in the final calculation. The most common mistake was omission of the factor 40/39 in finding the unbiased estimate of the population variance. Those who did not quote the z-value 2.576 explicitly, using instead, for example, non-standard notation such as `inv.norm(0.995)`, lost extra marks if their answer was wrong.

## Question 7 (b)

- (b) The secretary discovers that over a long period the value of  $\sigma^2$  is in fact 10.0. The secretary collects an independent random sample of 50 observations of  $T$  and constructs a new 99% confidence interval for  $\mu$  based on this sample of size 50, but using  $\sigma^2 = 10.0$ .

Find the probability that this new confidence interval contains the value  $\mu + 1.6$ .

[3]

Most candidates wrote their new confidence interval in the form  $\mu \pm 2.576\sqrt{0.2}$ , or even  $9.9 \pm 2.576\sqrt{0.2}$ , instead of the necessary  $\bar{X} \pm 2.576\sqrt{0.2}$ . This meant that they had no random variable in their inequality and were unable to make further progress.

## Misconception



Confidence intervals are centred on the mean of a particular *sample*, say  $\bar{x}$ , and not on the unknown population mean  $\mu$ .



## Question 8 (a)

- 8 A team of researchers have reason to believe that the number of calls received in randomly chosen 10-minute intervals to a call centre can be well modelled by a Poisson distribution. To test this belief the researchers record the number of telephone calls received in 60 randomly chosen 10-minute intervals. The results, together with relevant calculations, are shown in the following table.

							Total
Number of calls, $r$	0	1	2	3	4	$\geq 5$	
Observed frequency, $f$	18	13	12	9	8	0	60
$rf$	0	13	24	27	32	0	96
$r^2f$	0	13	48	81	128	0	270
Expected frequency	12.114	19.382	15.506	8.270	3.308	1.421	60
Contribution to test statistic	2.860	2.101	0.793	1.232			6.99

- (a) Calculate the mean of the observed number of calls received. [1]

This was almost always correct.

## Question 8 (b)

- (b) Calculate the variance of the observed number of calls received. [1]

This was almost always correct. A few candidates multiplied by  $60/59$ , which would give an unbiased estimate of the population mean, which is not what is asked for here.

## Question 8 (c)

- (c) Comment on what your answers to parts (a) and (b) suggest about the proposed model. [1]

This question illuminated a common distinction between candidates. Some candidates appreciated the difference between sample mean/variance and expected mean/variance. For a Poisson distribution, the *expected* mean and variance are equal, but parts (a) and (b) of this question give the *sample* mean and variance. Values from a sample will rarely match expected values exactly, and so the issue is not whether the answers to parts (a) and (b) are *equal* (many candidates wrote ' $1.6 \neq 1.94$ '), but whether their values are *similar*.

### Misconception



In an informal check for the validity of a model, it is not necessary for the sample mean and variance to be exactly equal to the expected values, or to satisfy exactly a theoretical equation. All that matters is that they should be close. Sample and expected values need to be carefully distinguished.

### Question 8 (d)

(d) Explain why it is necessary to combine some cells in the table. [1]

Careful candidates identified the cells where the expected frequencies are less than 5. Less careful ones said 'some cells have expected frequencies less than 5' or made general statements such as 'all the expected frequencies have to be greater than 5'.

### Question 8 (e)

(e) Show how the values 15.506 and 0.793 in the table were obtained. [4]

Most candidates could answer this fairly well, but in order to obtain full marks on this 'Show' question it was necessary not just to quote the probability 0.2584 (which could have been obtained from working backwards) but to explain where it came from. Best was to use the Poisson probability formula, but candidates could also gain credit by using a calculator value provided they gave enough significant figures to prove that they had found it themselves. Most justified the value 0.793 correctly.

### Question 8 (f)

(f) Carry out the test, at the 5% significance level. [5]

The question refers to 'a Poisson distribution', and not to the specific distribution  $Po(1.6)$ , so hypotheses should not refer to the parameter 1.6. Also, as 1.6 has been estimated from the data, a further degree of freedom has to be subtracted, so that the critical value is 5.991.

Some candidates did not see that the test statistic, 6.99, was given in the table and calculated it (sometimes wrongly). They lost time, but not marks.

Some less successful responses saw candidates give the hypotheses the wrong way round, or subtract only 1 from the number of cells, thus obtaining a totally wrong value for the number of degrees of freedom. Some tried to use the formula  $(n - 1)(m - 1)$ .

### Question 8 (g)

In the light of the result of the test, the team consider that a different model is appropriate. They propose the following improved model:

$$P(R = r) = \begin{cases} \frac{1}{60}(a + (2-r)b) & r = 0, 1, 2, 3, 4 \\ 0 & \text{otherwise,} \end{cases}$$

where  $a$  and  $b$  are integers.

- (g) Use **at least three** of the observed frequencies to suggest appropriate values for  $a$  and  $b$ . You should consider more than one possible pair of values, and explain which pair of values you consider best. (Do not carry out a goodness-of-fit test.) **[3]**

Many candidates were able to make some progress with this test of modelling skills. Some did not see that the observed frequencies had to be divided by 60, and some used the expected frequencies instead. Many candidates were able to make a reasoned choice between apparently inconsistent final answers.

It is necessary for  $a$  to be 12 because otherwise the probabilities do not add up to 1. Several possible values of  $b$  could obtain full credit.

A few candidates wrongly 'recognised' the format of the expression for  $P(R = r)$ , with its curly bracket and the word 'otherwise', and tried to use integration, as if this were the PDF of a continuous random variable. This is not an appropriate way to recognise a method.

---

# 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.



OCR is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored. © OCR 2023 Oxford Cambridge and RSA Examinations is a Company Limited by Guarantee. Registered in England. Registered office The Triangle Building, Shaftesbury Road, Cambridge, CB2 8EA. Registered company number 3484466. OCR is an exempt charity.

OCR operates academic and vocational qualifications regulated by Ofqual, Qualifications Wales and CCEA as listed in their qualifications registers including A Levels, GCSEs, Cambridge Technicals and Cambridge Nationals.

OCR provides resources to help you deliver our qualifications. These resources do not represent any particular teaching method we expect you to use. We update our resources regularly and aim to make sure content is accurate but please check the OCR website so that you have the most up to date version. OCR cannot be held responsible for any errors or omissions in these resources.

Though we make every effort to check our resources, there may be contradictions between published support and the specification, so it is important that you always use information in the latest specification. We indicate any specification changes within the document itself, change the version number and provide a summary of the changes. If you do notice a discrepancy between the specification and a resource, please [contact us](#).

You can copy and distribute this resource freely if you keep the OCR logo and this small print intact and you acknowledge OCR as the originator of the resource.

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

Whether you already offer OCR qualifications, are new to OCR or are thinking about switching, you can request more information using our [Expression of Interest form](#).

Please [get in touch](#) if you want to discuss the accessibility of resources we offer to support you in delivering our qualifications.