Qualification Accredited

Oxford Cambridge and RSA

AS LEVEL

Examiners' report

FURTHER MATHEMATICS B (MEI)

H635

For first teaching in 2017

Y412/01 Summer 2024 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 that 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 Y412/01 series overview

This is a calculator paper, in which situations are modelled by discrete random variables and the suitability of models is tested using chi-squared tests. Bivariate data are investigated, with tests for correlation and association, as well as modelling using regression.

To do well on this paper, candidates must demonstrate their understanding of the content from the specification. Candidates should be able to choose and apply appropriate models and be able to provide structured responses to questions involving hypothesis tests.

Once again, the overall performance was good this year. It was pleasing to see suitable structure in hypothesis tests with appropriate, non-assertive conclusions provided. It was pleasing that candidates could identify and use appropriate probability distributions.

Candidates should be aware that questions asking for 'detailed reasoning' to be given, or using command words such as 'show that' or 'determine', must include relevant supporting working for their final answer.

Candidates could have performed better in answering the questions where explanation or justification was required.

Candidates who did well on this paper generally:	Candidates who did less well on this paper generally:
• chose and quoted an appropriate probability distribution and stated its parameters • provided working to show how answers were calculated • showed good understanding of different probability distributions • provided concise, focused explanations and used context when needed • showed some understanding of residuals • provided correct answers to the questions involving probability • showed good understanding of hypothesis testing.	• chose inappropriate probability distributions • provided little if any working to support answers • were unable to distinguish between different probability distributions • gave explanations of a general nature rather than focusing on statistical features • showed little understanding of residuals • had limited success with probability calculations that involved a range of values • showed limited understanding of hypothesis testing.

Question 1 (a)

1 The probability distribution for a discrete random variable X is given in the table below.

x	0	1	2	3
$P(X = x)$	$2c$	$3c$	$0.5 - c$	c

(a) Find the value of c. **[2]**

This was answered well by most candidates. The few errors that occurred were from solving the equation $0.5 + 5c = 1$ incorrectly.

Question 1 (b)

(b) Find the value of each of the following.

- $E(X)$
- $Var(X)$ **[3]**

Some candidates used $Var(X) = E(X^2)$.

Question 1 (c)

The random variable Y is defined by $Y = 2X - 3$.

(c) Find the value of each of the following.

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

Many candidates did not use $Var(Y) = 2^2 \times Var(X)$.

Question 2 (a)

2 In a game of chance there are 32 slots, numbered 1 to 32, and on each turn a ball lands in one of them. You may assume that the process is completely random.

You are given that X is the random variable denoting the number of the slot that the ball lands in on a given turn.

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

Most candidates recognised this as a uniform distribution, but many did not succeed in identifying it fully as a discrete uniform distribution on the values 1, 2, …, 32. Some candidates thought that a geometric distribution was required.

Question 2 (b)

(b) Write down $\mathrm{P}(X = 7)$. **[1]**

Most candidates provided a correct answer here. Some of the candidates that in part (a) responded with 'geometric distribution' managed to recover here, although 0.0258 was a common incorrect answer.

Question 2 (c)

Players of the game start with a score of 0. On each turn a player may choose to play the game by selecting a number. If the ball lands in the slot with that number then 15 is added to the player's score. Otherwise, the player's score is reduced by 1. A player's score may become negative.

A player decides to play the game, selecting the number 7 on each turn, until the ball lands in the slot numbered 7.

You are given that Y is the random variable denoting the number of turns up to and including the turn in which the ball lands in the slot numbered 7.

(c) Determine $\mathrm{P}(Y \leqslant 15)$. **[3]**

Many candidates provided a correct answer here, although not all were supported with a probability calculation (see below).

Exemplar 1

$Y \sim Geo\left(\frac{1}{32}\right)$

$P(Y \leq 15) = 1 - P(X > 15)$

$= 1 - \left(\frac{31}{32}\right)^{15}$

$= 1 - 0.6211203132$

$= 0.3788796868 = \boxed{0.3789 \text{ (4sf)}}$

Assessment for learning

The command word 'determine' indicates that answers must be supported with appropriate working. In this instance, candidates should have stated the distribution and indicated the method used to arrive at the final answer (as seen in Exemplar 1).

Question 2 (d)

(d) Determine the player's expected final score. **[3]**

Many of those who found the expected number of turns as 32 did not get further. Of those who did, many provided the correct final answer with a supporting calculation, but some did not allow for the final turn being included among the 32 expected turns.

Question 3 (a)

3 A glassware factory produces a large number of ornaments each week. Just before they leave the factory, all the ornaments are checked and some may be found to be defective. The Quality Assurance Manager of the factory wishes to model the number of defective ornaments that are found each week using a Poisson distribution.

The numbers of defective ornaments found each week in a period of 40 weeks are shown in **Table 3.1**.

Table 3.1

No. of defective ornaments in a week, r	0	1	2	3	4	5	6	$\geqslant 7$
No. of weeks with r defective ornaments, f	2	14	13	5	3	1	2	0

You are given that summary statistics for the data are $\sum f = 40$, $\sum rf = 84$ and $\sum r^2 f = 256$.

(a) By using the summary statistics to determine estimates for the mean and variance of the number of defective ornaments produced by the factory each week, explain how the data support the suggestion that the number of defective ornaments produced each week can be modelled using a Poisson distribution. **[3]**

Most candidates obtained the correct value for the sample mean and showed awareness that in a Poisson distribution the mean is equal to the variance. Few candidates used the sample variance ($\frac{1}{n-1}S_{rr}$), with most opting to use $\frac{1}{n}S_{rr}$.

Question 3 (b) (i)

The Quality Assurance Manager is asked by the head office to carry out a chi-squared hypothesis test for goodness of fit based on a Po(2) distribution.

(b) **Table 3.2**, which is incomplete, gives observed frequency, probability, expected frequency and chi-squared contribution.

Table 3.2

No. of defective ornaments in a week, r	Observed frequency	Probability	Expected frequency	Chi-squared contribution
0	2	0.13534	5.4134	2.15232
1	14			
2	13	0.27067		0.43620
3	5		7.2179	
$\geqslant 4$	6	0.14288		0.01421

(i) Complete the copy of the table in the Printed Answer Booklet. **[4]**

This was well answered by most, with suitably accurate values provided.

Question 3 (b) (ii)

(ii) Carry out the test at the 10% significance level. **[6]**

This proved to be challenging for many. Some candidates struggled to provide suitable wording in their hypotheses (of these, some thought that the data was the model, not the Poisson distribution). Many candidates interchanged the hypotheses. In general, candidates struggled to identify the correct number of degrees of freedom (3 degrees of freedom was a common incorrect choice). Most showed a suitable comparison of test statistic with a critical value, accompanied by an appropriate decision. It was encouraging to see most candidates provide a non-assertive conclusion.

Question 3 (c)

(c) On one occasion a fork-lift truck in the factory drops a crate containing eight ornaments and all of them are subsequently found to be defective.

Explain why the Poisson model cannot model defects occurring in this manner. **[1]**

Although many candidates earned this mark, many of them provided a list of plausible reasons that did not all relate to a Poisson distribution. Better responses included reference to 'defects' and focused on a key issue, such as the defects not occurring independently or defects not occurring singly.

Assessment for learning

Candidates should avoid making general comments that have no relevance to either the context or the probability model.

In addition to learning the key features of the probability distributions that appear in this specification, candidates should also consider the similarities and differences between these distributions; this should help them provide relevant and concise explanations in context.

Question 4 (a)

4 A chemist is conducting an experiment in which the concentration of a certain chemical, A, is supposed to be recorded at the start of the experiment and then every 30 seconds after the start. The time after the start is denoted by t s and the concentration by z mg cm^{-3}. The collected data are shown in the table below. Note that the concentration at $t = 90$ was not recorded.

Time, t	0	30	60	120	150
Concentration of A, z	40.0	31.3	27.5	12.8	11.4

The chemist wishes to plot the data on a graph.

(a) Explain why t should be plotted on the horizontal axis. **[1]**

Most candidates provided a suitable explanation here. Leeway was given for those who did not use a correct term, provided that the meaning was clear. Candidates are advised to use statistical terms in their explanations, such as those mentioned in the specification.

Question 4 (b) (i)

You are given that the summary statistics for the data are as follows.

$n = 5 \quad \sum t = 360 \quad \sum z = 123.0 \quad \sum t^2 = 41400 \quad \sum z^2 = 3629.74 \quad \sum tz = 5835$

The regression line of z on t is given by $z = a + bt$ and is used to model the concentration of chemical A for $t \geqslant 0$.

(b) (i) Use the summary statistics to determine the value of a and the value of b. **[4]**

Overall, this was answered well, but some candidates mixed up their *a* and *b* values. Candidates were expected to show working to support their answers here and most did so.

Question 4 (b) (ii)

(ii) Find the value of the residual at each of the following values of t.

- $t = 60$
- $t = 120$ **[2]**

This was not well answered. Many candidates obtained just predictions rather than residuals for each value of *t*. Many of those who attempted to calculate residuals either gave two positive values, or used 'residual = predicted value – observed value'.

Question 4 (c) (i)

(c) (i) Use the equation of the regression line to estimate the value of the concentration at 90 seconds. **[1]**

As the values provided in the question for concentration, *A*, were given to just 3 significant figures, final answers given to 4 or more significant figures were overspecification and so were penalised here. Most candidates gave answers between 4 and 9 significant figures.

Question 4 (c) (ii)

(ii) With reference to your answers to part **(b)(ii)**, comment on the reliability of your answer to part **(c)(i)**. **[1]**

Many candidates did not follow the instruction to respond with 'reference to your answers to part **(b)(ii)**'. Many simply referred to interpolation and did not include any reference to the residuals found earlier.

Assessment for learning

Candidates would benefit from learning how to calculate residuals and interpret their values.

Question 4 (d)

Further experiments indicate that the model is reasonably reliable for times greater than 150 seconds up to about 200 seconds.

(d) Show that the model **cannot** be valid beyond a time of about 200 seconds. **[1]**

This was answered reasonably well; however, many candidates did not use their regression line equation to provide a suitable value to back up their statements.

Question 5 (a)

5 A student is investigating possible association between the amount of coffee that an adult drinks each day and the number of hours that they remain awake each day. In an initial investigation, a random sample of 8 adults is selected. The student obtains the following information from each of these adults: the amount of coffee that they drink each day and the number of hours that they remain awake each day.

The student analyses the data and finds that the associated product moment correlation coefficient is 0.6030.

(a) State **one** assumption that must be made for a hypothesis test based on the product moment correlation coefficient to be carried out. **[1]**

Many candidates did not show awareness of the distinction between 'data in a sample' and 'population from which the data has been drawn'. Incorrect answers such as the one shown in Exemplar 2 below were common (the candidate in Exemplar 2 has also muddled their reference to a 'bivariate normal distribution').

Exemplar 2

5(a)	the data must be bivariate and normally distributed.

Misconception

Candidates need to know the distinction between 'data in a sample' and 'population from which the data has been drawn'. In questions such as this, candidates should state the modelling assumption that the data are drawn from a bivariate normal distribution (or, alternatively, the modelling assumption that the underlying population has a bivariate normal distribution).

Question 5 (b)

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

(b) Carry out a test at the 5% significance level to investigate whether there is any correlation between amount of coffee drunk and number of hours awake. **[5]**

Most candidates performed well here. A common error was not to refer to population, either when defining the symbol ρ or when giving hypotheses in words. Some of those who used ρ in their hypotheses did not define it contextually. Most candidates identified the correct critical value and then proceeded to make a suitably non-assertive conclusion based on a correct comparison with the test statistic.

Question 5 (c)

The student conducts a second investigation which is similar to the first but this time based on a random sample of 30 adults. The product moment correlation coefficient for the new data is 0.5487. The student carries out an equivalent hypothesis test to the one carried out in part **(b)**, again using a 5% significance level.

(c) Identify any **differences** between the two tests and their results. You do **not** need to restate the hypotheses or explain the conclusion in context. **[2]**

Many candidates stated both the different critical value and different result for this second test. Despite the advice given, many candidates referred to the context.

Question 5 (d)

(d) You may assume the following guidelines for considering effect size.

Product moment correlation coefficient	Effect size
0.1	Small
0.3	Medium
0.5	Large

Explain briefly why the results of the student's second investigation are likely to be more reliable than the results of the initial investigation. **[2]**

Most candidates earned a mark for noting that the larger sample in the second investigation increased the likelihood that its results would be more reliable than for the first investigation.

Very few candidates provided adequate explanations that referred to effect size. The large effect size for the second (larger) sample indicated that the significant result has some practical significance.

Question 6 (a) (i)

6 A bank monitors the amounts of cash withdrawn from a cash machine. It categorises any withdrawal of an amount of £50 or less as 'small' and any withdrawal of an amount greater than £50 as 'large'.

Over a long period of time the bank finds that the proportion of withdrawals that are small is 0.43.

The bank wishes to model a sample of 10 withdrawals to examine the number of small withdrawals.

(a) **(i)** State a suitable probability distribution for such a model, justifying your answer. **[3]**

Most candidates identified that a binomial distribution is suitable in this situation and many of these also provided the correct parameters. Attempts to justify the answer were less successful however, either missing one of the key features or including features relating to other distributions.

Question 6 (a) (ii)

(ii) State **one** assumption needed for the model to be valid. **[1]**

From the answers given, it seemed that many candidates were guessing here. Those who had a good understanding of the requirements for using a binomial model were successful in providing concise answers in context. Most successful responses commented that withdrawals need to be independent; fewer commented that the sample of withdrawals must be selected randomly.

Too many candidates were not sufficiently specific. Saying that outcomes/events must be independent was deemed insufficient. Those who stated that the probability of each withdrawal must be independent did not earn the mark.

Question 6 (b) (i)

(b) **(i)** Find the probability that exactly 4 of the 10 withdrawals are small. **[1]**

Most candidates earned this mark. Overspecification was common, but not penalised.

Question 6 (b) (ii)

(ii) Find the probability that exactly 4 of the 10 withdrawals are large. **[1]**

Most candidates earned this mark. Overspecification was common, but not penalised.

Question 6 (b) (iii)

(iii) Find the probability that no more than 4 of the 10 withdrawals are large. **[2]**

Stronger responses included reference to the probability distribution being used, as well as an indication of finding $P(X \geq 6)$ where X represents the number of small withdrawals.

Question 6 (c)

(c) Find the probability that, in the 10 withdrawals, the 7th withdrawal is large and there are exactly 3 that are small. **[2]**

Very few responses showed understanding of what was required. Most of the successful responses adopted the method outlined in the mark scheme, however a few used an equivalent calculation based on the B(10, 0.43) distribution.

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.

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