Qualification Accredited

Oxford Cambridge and RSA

A LEVEL

Examiners' report

FURTHER MATHEMATICS B (MEI)

H645
For first teaching in 2017

Y432/01 Summer 2019 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. 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 can be downloaded from OCR.

Paper Y432 series overview

This was the first sitting of this Statistics Minor (Y432/01) component for the new revised A Level examination for GCE Further Mathematics B (MEI). It is an optional examined component contributing one third of the total mark for the applied part of the course.

To do well on this paper candidates needed to be able to:

- Identify appropriate probability distributions to use in a variety of different problems.
- Carry out hypothesis tests using correct terminology and stating appropriate non-assertive conclusions in context.
- Discuss features of different sampling methods, and show understanding of the assumptions necessary to use distributions and carry out tests.
- Efficiently use calculators that are able to calculate means, standard deviations, correlation coefficients, equations of regression lines and probabilities, including cumulative probabilities, from the binomial and Poisson distributions.
- Choose an appropriate level of accuracy to give answers to. Calculations within a probability model should use exact numbers (e.g. fractions) where possible or decimal numbers to four decimal places.

Candidates who did well on this paper were able to successfully identify the appropriate calculations necessary to solve a variety of problems using statistics and probability. They were also able to discuss the merits, and limitations, of a variety of sampling and testing methods. When drawing conclusions and interpreting results they were able to do so in a suitably nuanced fashion.

Candidates who did less well were able to accurately carry out a number of calculations. These included probabilities, chi-squared expected values and test statistic and a product moment correlation coefficient.

Generally candidates performed very well on questions involving numerical calculations. Good understanding was shown when choosing distributions, calculating probabilities and various statistics.

Candidates fared less well on questions requiring written explanations.

Question 1 (a)

1 In a game at a charity fair, a spinner is spun 4 times.
On each spin the chance that the spinner lands on a score of 5 is 0.2.
The random variable X represents the number of spins on which the spinner lands on a score of 5.

(a) Find $\mathrm{P}(X = 3)$. **[2]**

The majority of candidates answered this correctly.

Question 1 (b)

(b) Find each of the following.

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

Most candidates successfully used *np* and *np*(1 – *p*). Some produced a probability table and calculated the values which was time consuming and more likely to result in a numerical error.

Question 1 (c) (i)

One game costs £1 to play and, for each spin that lands on a score of 5, the player receives 50 pence.

(c) **(i)** Find the expected total amount of money gained by a player in one game. **[2]**

Usually answered well with the loss being indicated either in words or by a negative value. Lower ability candidates overlooked the £1 cost of the game.

Question 1 (c) (ii)

(ii) Find the standard deviation of the total amount of money gained by a player in one game. **[1]**

This part was usually answered well. Common errors were to leave the final answer as the variance or not to give the answer in pounds or pence.

Question 2 (a)

2 A market researcher wants to interview people who watched a particular television programme. Audience research data used by the broadcaster indicates that 12% of the adult population watched this programme. This figure is used to model the situation.
The researcher asks people in a shopping centre, one at a time, if they watched the programme. You should assume that these people form a random sample of the adult population.

(a) Find the probability that the fifth person the researcher asks is the first to have watched the programme. **[2]**

Most candidates used the correct approach. Some candidates only gave the value to three decimal places and so lost a mark.

Question 2 (b)

(b) Find the probability that the researcher has to ask at least 10 people in order to find one who watched the programme. **[1]**

Usually answered successfully, sometimes by calculating the complement first. A number of candidates made the mistake of raising 0.88 to the power of 10.

Question 2 (c)

(c) Find the probability that the twentieth person the researcher asks is the third to have watched the programme. **[3]**

Higher ability candidates set their working out clearly. They broke the problem down by first considering the probability of two of the first 19 people watching the programme.

Exemplar 1

2(c) $0.88^{17} \times 0.12^{3} = 1.9667500027 \times 10^{-4}$

$= 1.967 \times 10^{-4}$

This exemplar illustrates a fairly common error which is to overlook the fact that the first two 'successes' could occur anywhere in the first 19.

Question 2 (d)

(d) Find how many people the researcher would have to ask to ensure that there is a probability of at least 0.95 that at least one of them watched the programme. **[3]**

The most common approach was to use logarithms. Most candidates using this approach correctly realised that they needed to give the integer above the decimal value even if they had run into difficulties with inequality signs in their solution. An equally valid approach was to use trial and improvement. For either method it was necessary to show sufficiently detailed working to justify the answer.

Question 3 (a)

3 A company has been commissioned to make 50 very expensive titanium components.
A sample of the components needs to be tested to ensure that they are sufficiently strong. However, this is a test to destruction, so the components which are tested can no longer be used.

(a) Explain why it would not be appropriate to use a census in these circumstances. **[1]**

Most candidates realised that using a census with a test to destruction would leave no usable components. Candidates commenting only about the expense of testing all of the items were not credited with the mark.

Question 3 (b)

A manager suggests that the first 5 components to be manufactured should be tested.

(b) Explain why this would not be a sensible method of selecting the sample. **[1]**

The best responses to this part related the comment to the context of the problem and noted how a change may occur in the production process.

Question 3 (c)

A statistician advises the manager that the sample selected should be a random sample.

(c) Give two desirable features (other than randomness) that the sample should have. **[2]**

Good answers to this question closely related comments to features of the sample rather than of the components. Some incorrect comments did not describe other features but elaborated on what a random sample was.

?	**Misconception**	It is not always desirable to have a large sample size as the context here demonstrates.

Question 4 (a)

4 Zara uses a metal detector to search for coins on a beach.
She wonders if the numbers of coins that she finds in an area of 10m^2 can be modelled by a Poisson distribution. The table below shows the numbers of coins that she finds in randomly chosen areas of 10m^2 over a period of months.

Number of coins found	0	1	2	3	4	5	6	>6
Frequency	13	28	30	14	10	2	3	0

(a) Software gives the sample mean as 1.98 and the sample standard deviation as 1.4212.
Explain how these values suggest that a Poisson distribution may be an appropriate model for the numbers of coins found. **[2]**

Most candidates showed that they understood the mean and variance should be similar and that they needed to make specific reference to the values in this question. Lower ability candidates either did not give any values or compared the mean and standard deviation.

Question 4 (b)

Zara decides to carry out a chi-squared test to investigate whether a Poisson distribution is an appropriate model.
Fig. 4 is a screenshot showing part of the spreadsheet used to analyse the data. Some values in the spreadsheet have been deliberately omitted.

	A	B	C	D
1	Number of coins found	Observed frequency	Expected frequency	Chi-squared contribution
2	0	13	13.8069	0.0472
3	1	28		
4	2	30	27.0643	0.3184
5	3	14	17.8625	0.8352
6	4	10	8.8419	0.1517
7	$\geqslant 5$	5		0.0015

Fig. 4

(b) Showing your calculations, find the missing values in each of the following cells.

- C3
- C7
- D3 **[4]**

This question was often answered very well. There was clear working in evidence showing appropriate use of calculator probability distribution functions. Most candidates understood the importance of giving values to four decimal places in keeping with the values in the table.

Question 4 (c)

(c) Explain why the numbers for 5, 6 and more than 6 coins found have been combined into the single category of at least 5 coins found, as shown in the spreadsheet. **[1]**

A significant number of candidates did not refer to the 'expected frequencies' in their answers. Comments often just mentioned 'frequencies' or 'observed frequencies'.

Question 4 (d)

(d) Complete the hypothesis test at the 5% level of significance. **[6]**

Most candidates did well to state their hypotheses at the start of their response, usually using an appropriate form of wording. A common error was to have the wrong number of degrees of freedom (usually 5). There were also a number of candidates who gave a critical value from the lower tail of the chi-squared table.

Candidates did well when comparing the test statistic and critical value to make a decision about the null hypothesis. In the written conclusion higher ability candidates took care to refer to the context in terms of the number of coins found.

Question 4 (e)

For the rest of this question, you should assume that the number of coins that Zara finds in an area of 10m^2 can be modelled by a Poisson distribution with mean 1.98.
Zara also finds pieces of jewellery independently of the coins she finds. The number of pieces of jewellery that she finds per 10m^2 area is modelled by a Poisson distribution with mean 0.42.

(e) Find the probability that Zara finds a total of exactly 3 items (coins and/or jewellery) in an area of 10m^2. **[2]**

Most candidates successfully combined the distributions for the coins and jewellery. The probability was sometimes only given to three decimal places and such responses were not credited the second mark. Some students attempted to consider various combinations of coins and jewellery, calculating a sum of products of probabilities. While possible to score full marks, this way was time consuming and more likely to lead to numerical errors.

Question 4 (f)

(f) Find the probability that Zara finds a total of at least 30 items (coins and/or jewellery) in an area of 100m^2. **[2]**

Most candidates did well making efficient use of calculators. Some used 30 instead of 29 when calculating the probability of less than 30.

Question 5 (a)

5 A student wants to know if there is a positive correlation between the amounts of two pollutants, sulphur dioxide and PM10 particulates, on different days in the area of London in which he lives; these amounts, measured in suitable units, are denoted by s and p respectively.
He uses a government website to obtain data for a random sample of 15 days on which the amounts of these pollutants were measured simultaneously. Fig. 5.1 is a scatter diagram showing the data.
Summary statistics for these 15 values of s and p are as follows.

$\Sigma s = 155.4 \qquad \Sigma p = 518.9 \qquad \Sigma s^2 = 2322.7 \qquad \Sigma p^2 = 21270.5 \qquad \Sigma sp = 6009.1$

Item removed due to third party copyright restrictions

Fig. 5.1

(a) Explain why the student might come to the conclusion that a test based on Pearson's product moment correlation coefficient may be valid. **[2]**

The best responses noted the rough shape of the scatter taking into account the outlier. Higher ability candidates were clear in their comments about what this suggested in terms of the underlying population distribution rather than the distribution of the data in the diagram.

Question 5 (b)

(b) Find the value of Pearson's product moment correlation coefficient. **[4]**

Most candidates carried out the calculation accurately and showed a suitable level of detail in their working out to make sure method marks were obtained in the event of a numerical error.

Question 5 (c)

(c) Carry out a test at the 5% significance level to investigate whether there is positive correlation between the amounts of sulphur dioxide and PM10 particulates. **[5]**

Most candidates clearly stated the hypotheses. Higher ability candidates defined the parameter in context. Some lower ability candidates used r as the parameter and were not credited.

The critical value was usually correctly identified; although some candidates mistakenly found their value from the Spearman's rank correlation table.

In this question the null hypothesis uses a specific population parameter value and the result is not significant. Higher ability candidates then showed an appreciation of the need to "not reject" rather than "accept" the null. They also put their conclusion in context.

Question 5 (d)

(d) Explain why the student made sure that the sample chosen was a random sample. **[2]**

Higher ability candidates recognised the fact that a random sample allows for a valid test and for inferences to be made about the population.

?	**Misconception**	Random samples are not always representative.

Question 5 (e)

The student also wishes to model the relationship between the amounts of nitrogen dioxide n and PM10 particulates p.
He takes a random sample of 54 values of the two variables, both measured at the same times. Fig. 5.2 is a scatter diagram which shows the data, together with the regression line of n on p, the equation of the regression line and the value of r^2.

Item removed due to third party copyright restrictions

Fig. 5.2

(e) Predict the value of n for $p = 150$. **[1]**

Answered successfully by nearly all candidates.

Question 5 (f)

(f) Discuss the reliability of your prediction in part **(e)**. **[2]**

Most candidates identified the relevance of interpolation. Higher ability candidates recognised that the scatter and/or the *r* squared value cast doubt on the reliability of the prediction.

Exemplar 2

5(f) It is reliable as 150 falls within the given range of data used for the regression line, ie it is not extrapolation.

This is a typical response which comments just on interpolation and is credited with 1 mark. No mention is made of the strength of the correlation either in terms of the diagram or the *r* squared value.

Question 6 (a)

6 The discrete random variable X has a uniform distribution over $\{n, n+1, ..., 2n\}$.

(a) Given that n is odd, find $P(X < \frac{3}{2}n)$. **[1]**

Higher ability candidates were able to write down the answer with no working. A successful approach from lower ability candidate's was to write down some lists for specific values of *n* to get a feel for the problem.

Question 6 (b)

(b) Given instead that n is even, find $P(X < \frac{3}{2}n)$, giving your answer as a single algebraic fraction. **[3]**

Writing lists was again a useful starting point for a number of candidates. It was possible to arrive at an apparently correct expression from incorrect reasoning (usually confusion between the value of *n* and the number of values that *X* could take lay at the heart of this). Higher ability candidates were able to give clear explanations to justify their algebra.

Question 6 (c)

(c) The sum of 6 independent values of X is denoted by Y.
Find $\text{Var}(Y)$. **[3]**

An attempt to use the formula for the variance of the discrete uniform distribution was regularly seen. Higher ability candidates understood the need to adapt this for (n + 1) possible values and also recognised that for a sum of independent values you just need to multiply by 6.

Exemplar 3

6(c)

$\mathrm{Var}(X) = \frac{1}{12}(n^2 - 1)$

$\mathrm{Var}(Y) = 36 \times \frac{1}{12} \times (n^2 - 1)$

$= 3n^2 - 3.$

This exemplar illustrates two common mistakes. Firstly the formula for the variance just uses *n* and secondly the candidate has not identified the difference between a sum of six independent values and multiplying one value by 6.

Supporting you

For further details of this qualification please visit the subject webpage.

Review of results

If any of your students' results are not as expected, you may wish to consider one of our review of results services. For full information about the options available visit the OCR website. If university places are at stake you may wish to consider priority service 2 reviews of marking which have an earlier deadline to ensure your reviews are processed in time for university applications.

activeresults

Review students' exam performance with our free online results analysis tool. Available for GCSE, A Level 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.

*To find out which reports are available for a specific subject, please visit ocr.org.uk/administration/support-and-tools/active-results/

Find out more at ocr.org.uk/activeresults

CPD Training

Attend one of our popular CPD courses to hear exam feedback directly from a senior assessor or drop in to an online Q&A session.

Please find details for all our courses on the relevant subject page on our website.

www.ocr.org.uk

The small print

Looking for a resource?

There is now a quick and easy search tool to help find **free** resources for your qualification:

www.ocr.org.uk/i-want-to/find-resources/

www.ocr.org.uk

OCR Customer Support Centre

General qualifications

Telephone 01223 553998
Facsimile 01223 552627
Email general.qualifications@ocr.org.uk

Cambridge Assessment

CERTIFIED ISO 9001 UKAS MANAGEMENT SYSTEMS 001