Qualification Accredited

Oxford Cambridge and RSA

AS LEVEL

Examiners' report

FURTHER MATHEMATICS A

H235

For first teaching in 2017

Y532/01 Summer 2024 series

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 responses 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 exam paper and the mark scheme can be downloaded from OCR.

Paper Y532/01 series overview

Candidates appeared well prepared for the more routine style questions, but many found questions involving more problem solving to be more challenging – for example question 4(c) and question 6(d). Overall, there was a good spread of results in the 30 to 50 range with a negative skew distribution and a few outlier results below the score of 15.

Unfortunately, a significant proportion of responses to questions requiring explanations saw candidates reproduce generic phrases that they had learnt instead of applying the concepts to the particular scenario defined in the question.

Many candidates seemed reluctant to make full use of algebra, preferring to use numerical values with mixed success.

Candidates who did well on this paper generally:	Candidates who did less well on this paper generally:
• were precise in their statements of hypotheses and conclusions to hypothesis tests • understood what is meant by 'modelling assumptions'. They could apply standard assumptions to the scenario of a question • were able to use algebra confidently in solving problems • were able to provide a summative assessment of the reliability of an estimate. They brought together several facts into a nuanced single conclusion.	• were vague or unspecific in their statements of hypotheses and conclusions • confused modelling assumptions with conditions or made only formulaic statements • were over-dependent on numerical calculations. They were unwilling to use algebra where necessary • were unable to coordinate several different facts in assessing outcomes.

Question 1 (a)

1 The random variable W can take values 1, 2 or 3 and has a discrete uniform distribution.

(a) Write down the value of $E(2W)$. **[1]**

The majority of candidates wrote down the correct answer to this part.

Question 1 (b)

(b) Find the value of $Var(2W)$. **[2]**

Although "Find" does not require any written working to be recorded, partial credit was available for those candidates that made an arithmetic mistake if they wrote down their method. Most candidates answered this correctly.

Question 1 (c)

(c) Determine the value of the constant k for which $E(2W+k) = Var(2W+k)$. **[2]**

A written method must be seen for "Determine" questions. This question was almost always answered correctly.

Question 1 (d)

The random variable S has the probability distribution shown in the following table.

s	2	3	4	5	6
$P(S = s)$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{2}{9}$

(d) Calculate $Var(S)$. **[3]**

This question was answered almost always correctly, although one or two candidates tried to use the formula for a uniform distribution.

Question 2 (a)

2 For a random sample of 160 employees of a large company, the principal method of transport for getting to work, arranged according to grade of employee, is shown in the table.

Grade	Walk or cycle	Private motorised transport	Public transport
A	9	13	6
B	16	43	41
C	11	8	13

A test is carried out at the 5% significance level of whether there is association between grade of employee and method of transport.

(a) State appropriate hypotheses for the test. **[1]**

This was generally well done. This question is about "association" and the word 'correlation' should not be used in this context.

Question 2 (b)

The contributions to the test statistic are shown in the following table, correct to 3 decimal places.

Grade	Walk or cycle	Private motorised transport	Public transport
A	1.157	0.289	1.929
B	1.878	0.225	0.327
C	2.006	1.800	0.083

(b) Show how the value 0.225 is obtained. **[3]**

This was generally well done, although some candidates did not provide sufficient detail to "Show" how the expected frequency was obtained.

Question 2 (c)

(c) Complete the test, stating the conclusion. **[3]**

The calculation and comparison were generally correct. The conclusion was often poorly stated. It is wrong to say, 'there is significant evidence that grade is dependent on method of transport', as it is not possible to infer that one factor depends on the other. Some attempted to use a double negative, such as 'there is insufficient evidence of no association', which would be appropriate if the null hypotheses were not rejected. Some wrote 'there is sufficient evidence that there may be association', but this is too vague; in any test it is always true that there may be association.

Question 2 (d)

(d) Which combination of grade of employee and method of transport most strongly suggests association? Justify your answer. **[1]**

There were many correct responses. However, many identified the cell with the *smallest* overall contribution. Some considered only one of the variables, such as 'Grade C'.

Question 3 (a)

3 The ages, x years, and the reaction time, t seconds, in an experiment carried out on a sample of 15 volunteers are summarised as follows.

$n = 15 \quad \sum x = 762 \quad \sum t = 8.7 \quad \sum x^2 = 44204 \quad \sum t^2 = 5.65 \quad \sum xt = 490.1$

(a) Calculate the value of the product moment correlation coefficient between x and t. **[2]**

This was generally correct. Some who did not get the correct answer could get one mark by showing appropriate intermediate working, such as correct values of at least two of S_{xx}, S_{tt} and S_{xt}.

Question 3 (b)

(b) Calculate the equation of the line of regression of t on x. Give your answer in the form $t = a + bx$ where a and b are constants to be determined. **[2]**

This was generally well answered. Most candidates gave the correct letters as well as the correct coefficients.

Question 3 (c)

(c) Explain the relevance of the quantity $\sum(t-a-bx)^2$ to your answer to part **(b)**. **[1]**

This question was rarely correct. Many said that it equalled 0 or gave the variance. Some said that it was the sum of squares of the discrepancies, but this does not address the question, which asked about its relevance to the calculated line of regression. The key point is that this quantity is minimised.

Question 3 (d)

(d) Estimate the reaction time, in seconds, for a volunteer aged 42. **[1]**

This question was generally correct.

Question 3 (e) (i)

It is subsequently decided to measure the reaction time in tenths of a second rather than in seconds (so, for example, a time of 0.6 seconds would now be recorded as 6).

(e) (i) State what effect, if any, this change would have on your answer to part **(a)**. **[1]**

This question was almost always correct, although many candidates confused 'effect' with 'affect'.

Question 3 (e) (ii)

(ii) State what effect, if any, this change would have on your answer to part **(b)**. **[1]**

There were far fewer correct responses here. The question asked about the effect on the equation, not on the answers, so 'The answers are multiplied by 10' does not meet the question demand. A response such as 'they would both increase' was insufficient. 'Both *a* and *b* are multiplied by 10' is correct and sufficient.

Question 3 (f)

It is known that the sample of 15 volunteers consisted almost entirely of students and retired people.

(f) Using this information, and the value of the product moment correlation coefficient, comment on the reliability of your estimate in part **(d)**. **[3]**

This question required candidates to make a nuanced assessment of the overall reliability of an estimate. They had to bring together two different pieces of information – a very common situation with statistics and one that represents a central part of the subject.

Candidates needed to:

- explain the consequences of most of the data not being close to 42,
- comment on the fact that $r = 0.836$ suggests strong correlation, and
- draw a single conclusion.

Many wrote 'the value of *r* suggests strong correlation, so the estimate is reliable, but 42 is not close to most of the data, so it is less reliable.' This does not give a single conclusion. The two facts need to be coordinated, although this conclusion might be indefinite, even 'we cannot tell'.

Some said that 'the value of *r* was unreliable', as in Exemplar 1 below. On its own this is inadequate, as the value of *r* is completely known. However, some insightful candidates said that you could easily have two clusters of results. One cluster might be younger people with short reaction times and the other cluster might be older people with longer reaction times. Within each cluster the variables might not be correlated, but the effect of the two clusters would still suggest a spurious overall correlation. It would be better to say that the value of *r* was misleading.

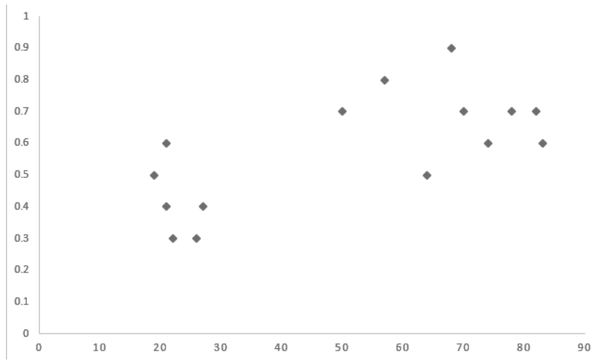
Exemplar 1

3(f) There is very little data available for working adults
∴ the value in part (d) is unreliable
The value of r is also unreliable and would probably
change with the addition of working adults.

This also shows two separate assessments of reliability, but not a single overall conclusion. The statement 'the value of *r* is unreliable' is imprecise. The value of the test statistic is unquestionably correct, as it is a calculation from given data; what is unreliable is the conclusion that can be drawn from it. 'The value of *r* may be misleading' would be a better statement.

Assessment for learning

Teachers might find it useful demonstrate this scenario to students by drawing scatter diagrams which corresponded to high values of *r* but where a straight line would not provide reliable estimates. Here is a chart similar to that for the data in the question.



Conclusions such as 'the estimate may be unreliable' or 'if there is some data close to 42 it would be reliable' are too non-committal. Any estimate may be unreliable, and candidates need to make an assessment on the basis of the available information, although 'there is not enough information to make an assessment' might be appropriate in some contexts.

Question 4 (a)

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

The random variables X and Y denote the number of telephone calls and the number of e-mails, respectively, received by a company in a randomly chosen one-minute period in a working day.

For any one-minute period, the following assumptions can be made.

- Telephone calls are received randomly and independently of one another.
- E-mails are received randomly and independently of one another.
- The average rate at which telephone calls are received is constant.
- The average rate at which e-mails are received is constant.
- Telephone calls and e-mails are received independently of one another.

It is known that $\mathrm{E}(X) = 3$.

(a) Find the probability that 4 telephone calls are received in a randomly chosen one-minute period. **[1]**

This question was almost always correct.

Question 4 (b)

(b) A sample of 10 independent observations of X is obtained.

Find the expected number of these 10 observations that are in the interval $2 < X < 8$. **[3]**

The correct response was seen quite often, but so was 7.89 from including 2 in the interval. Most realised that they needed simply to multiply the probability p by 10, but some attempted to work out p^{10} or similar.

Question 4 (c)

It is also known that

$$P(X + Y = 4) = \frac{27}{8}P(X = 2) \times P(Y = 2).$$

(c) Determine the possible values of $E(Y)$. **[7]**

Although this was an extended problem-solving question, there were quite a few correct responses. Less successful candidates tried to use $P(X + Y = 4) = P(X = 4) + P(Y = 4)$. The crucial step was the realisation that the distribution of $X + Y$ is $Po(3 + \lambda)$.

Candidates who were less confident with algebra often used a numerical value of $P(X = 2)$ and then often struggled later to get rid of the factor e^{-3} on the left-hand side.

Some candidates used the excellent short-cut of going from $4(3 + \lambda)^4 = 729\lambda^2$ to $2(3 + \lambda)^2 = \pm 27\lambda$. However, the majority of successful solutions multiplied the equation out, obtained a messy quartic, and then solved this by calculator. Whatever method was used, it was necessary to ignore negative signs at some stage, and this needed a few words of justification to reflect the "Determine" command word.

Question 4 (d)

(d) Explain where in your solution to part **(c)** you have used the assumption that telephone calls and e-mails are received independently of one another. **[1]**

The assumption is used in order to say that $X + Y$ has a Poisson distribution. Few candidates realised this. Many thought that it was used when the means are added, but this can be done even if the variables are not independent.

Question 5 (a)

5 In a fashion competition, two judges gave marks to a large number of contestants.

The value of Spearman's rank correlation coefficient, r_s, between the marks given to 7 randomly chosen contestants is $\frac{27}{28}$.

(a) An excerpt from the table of critical values of r_s is shown below.

Critical values of Spearman's rank correlation coefficient

	1-tail test	5%	2.5%	1%	0.5%
	2-tail test	10%	5%	2%	1%
n	6	0.8286	0.8857	0.9429	1.0000
	7	0.7143	0.7857	0.8929	0.9286
	8	0.6429	0.7381	0.8333	0.8810

Test whether there is evidence, at the 1% significance level, that the judges agree with each another. **[4]**

The statements of hypotheses and conclusions are made harder if candidates use the word 'agree', even though this is the wording in the question. The word 'agree' usually means 'positive correlation', as opposed to 'disagree' (negative correlation) and 'do not agree' (no correlation). It is better to use statements such as 'H_1: there is positive association between the judges' rankings', or better still to use symbols. Candidates who wrote 'H_0: the judges agree' tended to use a two-tailed test.

Question 5 (b)

The marks given by the two judges to the 7 randomly chosen contestants were as follows, where x is an integer.

Contestant	A	B	C	D	E	F	G
Judge 1	64	65	67	78	79	80	86
Judge 2	61	63	78	80	81	90	x

(b) Use the value $r_s = \frac{27}{28}$ to determine the range of possible values of x. **[4]**

Some candidates used the unranked data and were unable to make any progress. Most who started by showing that $\Sigma d^2 = 2$ were able to obtain the correct answer. The question required the use of the value of r_s, so solutions that did not use this value could not gain full credit even if the answer was correct.

Question 5 (c)

(c) Give a reason why it might be preferable to use the product moment correlation coefficient rather than Spearman's rank correlation coefficient in this context. **[1]**

This question had limited successful responses. Many responses indicated a lack of understanding of the logic of the question, for instance 'as the data is bivariate normal, Pearson's PMCC can be used'. There is nothing in the question to indicate that the data is bivariant normal. Good responses such as 'the PMCC uses all the data and not just the rankings', or 'to test for linear correlation rather than association' were quite common, but some said, 'because the data is ranked', which is not true.

Question 6 (a)

6 Anika walks along a street that contains parked cars. The number of cars that Anika passes, up to and including the first car that is white, is denoted by X.

(a) State **two** assumptions needed for X to be well modelled by a geometric distribution. **[2]**

As has happened in the past, a question about modelling assumptions was not well responded to by many candidates.

Misconception

In questions about modelling, the word 'assumptions' has a technical meaning and is not well understood. Assumptions are not conditions which allow you to select a particular distribution. For example, 'the number of repeated trials is fixed' and 'outcomes can be classified as success or fail' are conditions that tell you that a binomial distribution is likely to be considered. There is rarely any doubt about whether conditions are satisfied. Issues like 'cars must be either white or not' are merely a matter of making a specific decision when collecting data. They have nothing to do with modelling and are never appropriate in questions about assumptions. But once the distribution has been selected, assumptions are needed for the calculations to be valid, and usually there is doubt about these.

To summarise:

- Conditions tell you which type of distribution you are likely to use – binomial, geometric, Poisson, etc.
- Once this distribution has been selected, assumptions are needed to make calculations valid.

Exemplar 2

6(a) 1. the cars must be ~~travelling~~ parked independently of one another

The meanings of the assumptions in any particular context are often not understood. For example, in this scenario 'independence' means 'whether or not any one car is white is independent of whether any other car is white', and not, for example, the response in this exemplar.

The other assumption is most precisely stated as 'the probability that any particular car is white is constant' (there might, for example, be social reasons why white cars might be more popular in a particular part of the street).

Assessment for learning

It would be helpful for teachers to discuss the precise meanings of the standard assumptions in particular contexts. The focus should be on understanding, rather than merely learning generic phrases that can be memorised without understanding. Where consistent with candidates' overall needs, the aim should be a precise use of language and expression.

Question 6 (b)

Assume now that X can be well modelled by the distribution $\text{Geo}(p)$, where $0 < p < 1$.

(b) For $p = 0.1$, find $\text{P}(X > 6)$. **[2]**

Answers to this question were generally correct.

Question 6 (c)

The number of cars that Anika passes, up to **but not including** the first car that is white, is denoted by Y.

(c) For a general value of p, determine a simplified expression for $\mathrm{E}(Y) \div \mathrm{Var}(Y)$, in terms of p. **[3]**

Many candidates did not appreciate the effect of not counting the first car that was white. The effect is that $Y = X - 1$. Almost everyone correctly said that $\mathrm{Var}(Y) = \frac{1-p}{p^2}$, but many thought that E(Y) was $\frac{1}{p}$, whereas in fact it is $\frac{1}{p} - 1$.

Question 6 (d)

Ben walks along a different street that also contains parked cars. The number of cars that Ben passes, up to and including the first white car **on which the last digit of the number plate is even** is denoted by Z.

It may be assumed that Z can be well modelled by the distribution $\mathrm{Geo}(\frac{1}{2}p)$, where p is the parameter of the distribution of X.

It is given that $\mathrm{P}(Z = 3) = k\mathrm{P}(X = 3)$, where k is a positive constant.

(d) Determine the range of possible values of k. **[5]**

Some candidates assumed that p was still 0.1 and tried to use numerical values, but that is not what the structure of the question indicates. Most were able to write down the correct equation, but some made it very difficult for themselves by not cancelling the factor of p. Some obtained a quadratic equation for p and then tried to use $b^2 - 4ac$, but this leads nowhere: the issue is that p is between 0 and 1, not that the equation has a solution without this restriction. A few candidates successfully used an algebraic argument, but most of the successful attempts used a numerical approach, substituting $p = 0$ to get $k = ½$ and then observing that higher values of p gave higher values of k. On this occasion a formal justification that k is an increasing function of p was not required, but it is not difficult to construct an informal argument.

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