Qualification Accredited

Oxford Cambridge and RSA

AS LEVEL

Examiners' report

FURTHER MATHEMATICS A

H235

For first teaching in 2017

Y532/01 Summer 2022 series

Version 1

Contents

Introduction

Our examiners' reports are produced to offer constructive feedback on candidates' performance in the examinations. They provide useful guidance for future candidates.

The reports will include a general commentary on candidates' performance, identify technical aspects examined in the questions and highlight good performance and where performance could be improved. A selection of candidate answers are 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.

Advance Information for Summer 2022 assessments

To support student revision, advance information was published about the focus of exams for Summer 2022 assessments. Advance information was available for most GCSE, AS and A Level subjects, Core Maths, FSMQ, and Cambridge Nationals Information Technologies. You can find more information on our website.

Paper Y532/01 series overview

There were many good scripts for this unit.

The most important common mistakes were not giving the hypotheses in a correct form when testing a correlation coefficient (Question 2) and assessing the validity of inferences with too high a degree of certainty (Questions 1 (d), 5 (c) and 7 (b) (ii)). In questions such as 1 (d) and 5 (c), some candidates suggested that a model was unlikely to be valid if the value of the test statistic was not equal to the corresponding parameter, whereas of course you would rarely or never expect them to be exactly equal.

It is particularly pleasing to report that there were many correct answers to Question 6, parts (a) and (b), on selections and arrangements. In the past this has been found a difficult topic.

As usual, less successful responses struggled with questions that were not very similar to ones that had been set in the past. Often such candidates try to answer questions as if they were identical to ones they had seen before, and this generally obtains no credit.

Candidates who did well on this paper generally did the following:	Candidates who did less well on this paper generally did the following:
• They assessed outcomes in nuanced terms and were not over-definite. • They answered questions about arrangements where some objects were not next to each other using the methods of "gaps" (or "barriers"), and then using permutations or combinations. • They gave conclusions to hypothesis tests in context and without being too definite. • They made inferences about which set of results came from different experiments by quantitative arguments, identifying distributions and using parameters.	• They made over-definite statements such as "the statistician is right". • They answered questions about arrangements where some objects were not next to one another by improvising methods, or by multiplying probabilities. • They gave conclusions to hypothesis tests without referring to the context or in too definite a manner (e.g., "Do not reject H_0. The null hypothesis is correct."). • They made inferences about which set of results came from different experiments by using qualitative arguments.

Key point call out: Statements of hypotheses

Hypotheses for tests for correlation should be stated in terms of parameter values, and the parameter should be defined. For example, H_0: $\rho_s = 0$, H_1: $\rho_s > 0$, where ρ_s is the population value of Spearman's rank correlation coefficient between the results in the two races.

Key point call out: Assessing validity

In assessing whether models or inferences are likely to be valid, candidates should not be over-definite in their conclusions. For example, "The point is probably an anomaly because the observed and expected values are very different", and not "The point is an anomaly, as the observed and expected values are not equal".

Question 1 (a) (i)

1 A geography student chose a certain point in a stream and took measurements of the speed of flow, $v \text{ ms}^{-1}$, of water at various depths, d m, below the surface at that point. The results are shown in the table.

d	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
v	0.8	0.5	0.7	1.2	1.1	1.3	1.6	1.4	0.4

$n = 9 \quad \sum d = 2.7 \quad \sum v = 9.0 \quad \sum d^2 = 0.96 \quad \sum v^2 = 10.4 \quad \sum dv = 2.85$

(a) (i) Explain why d is an example of an independent, controlled variable. **[1]**

Many candidates could produce an adequate explanation.

Question 1 (a) (ii)

(ii) Use **two** relevant terms to describe the variable v in a similar way. **[1]**

Very few candidates knew the correct term "response (variable)".

Question 1 (b)

A statistician believes that the point (0.5, 0.4) may be an anomaly.

(b) Calculate the equation of the least squares regression line of v on d for all the points in the table **apart from** (0.5, 0.4). **[2]**

This was usually well done, although some candidates who had correctly subtracted the appropriate values of d, v, d^2, v^2 and dv from the given totals then used $n = 9$ instead of 8, so that the answer $v = 0.178 + 3.18d$ was seen several times. Most candidates were able to use the correct variable letters in their answer.

Question 1 (c)

(c) Use the equation of the line found in part **(b)** to estimate the value of v when $d = 0.5$. **[1]**

This caused no problems.

Question 1 (d)

(d) Use your answer to part **(c)** to comment on the statistician's belief. **[1]**

Here it was necessary both to say that the difference between 0.4 and their 1.69 was large, and to conclude that the statistician was probably right. Many candidates merely said that the observed and predicted values were not equal, or that the statistician was right, or that it was an anomaly.

Exemplar 1

1(d)	Statistician is correct. (0.5, 0.4) is an anomaly according to the regression line.

The statement "Statistician is correct" is over-assertive. All that can be said is that it is likely that the statistician is correct. Further, this candidate has not supplied evidence for the conclusion, such as "the difference between 0.4 and 1.69 is large, so ...".

Question 1 (e)

(e) Use the diagram in the Printed Answer Booklet (which does **not** illustrate the data in this question) to explain what is meant by "least squares regression line". **[2]**

Most could draw appropriate lines on the diagram, although a few showed "discrepancies" that were perpendicular to the regression line, or horizontal. Some knew exactly what to write, but there were many insufficiently precise answers, ranging from "it gives the line of best fit" to "it minimises the squares of the differences". Candidates needed to say something like "it minimises the sum of the squares of the vertical differences".

Question 2

2 Eight runners took part in two races. The positions in which the runners finished in the two races are shown in the table.

Runner	A	B	C	D	E	F	G	H
First race	3	1	5	6	2	8	7	4
Second race	4	3	8	7	2	5	6	1

Test at the 5% significance level whether those runners who do better in one race tend to do better in the other. **[7]**

Specimen papers have made it clear that hypotheses for correlation coefficients need to be stated in terms of a parameter (here ρ_s) and to define that symbol, here "the population rank correlation coefficient between the race positions" (the word "population" is essential). Those who gave imprecise verbal statements did not score full marks. In any case it was necessary to indicate that this was a one-tailed test, preferably using "H_1: $\rho_s > 0$", or at least using the word "positive" in a verbal statement of H_1.

The calculation was generally done well. Some obtained a critical value from the tables for the PMCC instead of from Spearman; the two tables are constructed from different assumptions, so the critical values are not interchangeable even though either method of obtaining the test statistic will give the same answer (if the data is given as rankings).

Conclusions were often well stated, though it is incorrect to infer that "there is evidence that there is no correlation between the rankings". Hypothesis tests <u>never</u> give positive evidence that the null hypothesis is true.

Exemplar 2

2

$\Sigma d^2 = 1 + 4 + 9 + 1 + 9 + 1 + 9 = 34$

$r_s \quad \frac{6(34)}{8(63)} = 0.40476$

$n = 8, \quad CV = 0.6429$

let ρ be the population correlation coefficient

H_0: there is no association between doing better in 1 race over the ~~other~~ other

H_1: there's an association between doing better in a race over the other

$0.6429 > 0.40476$

$\therefore$ accept H_0, there is sufficient evidence to suggest no association between doing better in one race over the other

This candidate has used a verbal statement of the hypotheses (rather than using a parameter such as rho-subscript-*s*), and so cannot score more than 1 out of 2 for the hypotheses. In fact, this candidate has not stated the alternative hypothesis correctly as the word "positive" is required to show that this is a one-tailed test, so scores 0/2 for the hypotheses.

Passing over the incorrect calculation of the rank correlation coefficient, the candidate has not gained the final A1 as the statement "there is sufficient evidence to suggest no association" is incorrect. Acceptance of the null hypothesis does not provide evidence that it is true; all that can be said is that there is insufficient evidence that it is not true. In such cases, a clumsy double-negative such as this is often required, although a good response might be "there is insufficient evidence of positive correlation between the results".

Question 3 (a)

3 A discrete random variable X has the following probability distribution.

x	1	2	3	4
$\mathrm{P}(X=x)$	p	0.31	0.3	p^2

(a) Determine the value of p. **[3]**

This was generally very well done. Almost everyone obtained $p = 0.3$, but to gain full marks (as many did), candidates needed not just to reject $p = -1.3$ but to give a reason, such as "p must be ≥ 0".

Question 3 (b)

(b) It is given that $\mathrm{E}(aX+b) = \mathrm{Var}(aX+b) = 23.19$, where a and b are positive constants. Determine the value of a and the value of b. **[6]**

There were many good answers seen to this question, although some calculated the variance wrongly and a few used $a^2 \times \mathrm{E}(X) = 23.19$, or $a^2 \times \mathrm{Var}(X) + b = 23.19$, rather than the correct $a^2 \times \mathrm{Var}(X) = 23.19$.

Question 4

4 A school pupil keeps a note of whether her journeys to school and from school are delayed. The results for a random sample of journeys are shown in the table.

	Direction of journey	
	To school	From school
Delayed	64	56
Not delayed	74	106

Test at the 10% significance level whether there is association between delays and the direction of the journey. **[7]**

This was generally well done. Most candidates stated the hypotheses correctly (a verbal statement is correct here) and were able to use Yates's Correction in the calculations, though some gave the Yates formula incorrectly (often omitting the modulus signs). Most gave their final conclusions in suitably non-assertive terms, using phrases such as "there is evidence that ...".

Question 5 (a)

5 The manager of an emergency response hotline believes that calls are made to the hotline independently and at constant average rate throughout the day. From a small random sample of the population, the manager finds that the mean number of calls made in a 1-hour period is 14.4.

Let R denote the number of calls made in a randomly chosen 1-hour period.

(a) Using evidence from the small sample, state a suitable distribution with which to model R. You should give the value(s) of any parameter(s). **[1]**

This was almost always correct.

Question 5 (b) (i)

(b) In this part of the question, use the distribution and value(s) of the parameter(s) from your answer to part **(a)**.

(i) Find $P(R > 20)$. **[2]**

Apart from the predictably large number of candidates who used P(> 20) = 1 – P(≤ 19), this was usually correct.

Question 5 (b) (ii)

(ii) Given that $P(R = r) > P(R = r+1)$, show algebraically that $r > 13.4$. **[2]**

This was often well done, although those candidates who tried to use logarithms were usually unsuccessful. Most showed sufficient working.

Question 5 (b) (iii)

(iii) Hence write down the mode of the distribution. **[1]**

Relatively few candidates realised that the mode had to be an integer. The answers 13.4 and 14.4 were common.

Question 5 (c)

The manager also finds, from records over many years, that the modal value of R is 10.

(c) Use this result to comment on the validity of the distribution used in part **(b)**. **[1]**

Few candidates gave a suitably nuanced answer to this. It was necessary to indicate both that the modes of 10 and 14 were not very close and hence that the model was probably not valid, but many said categorically that it was not valid, or gave as a reason that 10 was not equal to 14. Some tried to answer a question they had not seen before by attempting to answer one that they had, for instance by referring to modelling assumptions.

A few candidates attempted to measure the difference between the observed and expected modes in terms of the number of standard deviations, but unfortunately the distribution of the sample mode is very different from that of a single result, so the approach doesn't work. (A very simplistic spreadsheet simulation, based arbitrarily on samples of size 20, shows that, if $\lambda = 10$, the probability that the sample mode is 14 or more is about 0.03.)

Question 5 (d)

(d) Assume now that the type of distribution used in part **(b)** is valid. Find the range(s) of values of the parameter(s) of this distribution that would correspond to the modal value of R being 10. **[2]**

Many candidates could not see what was needed here. A common answer was to calculate the probability of 10.

Few could relate the previous calculations to the mode being 10, and there was a lot of guesswork.

Question 6 (a)

6 A teacher has 10 different mathematics books. Of these books, 5 are on Algebra, 3 are on Calculus and 2 are on Trigonometry.

The teacher chooses 5 of the books at random.

(a) Find the probability that 3 of the books are on Algebra. **[3]**

A considerable number of correct answers were seen to this question, which was very pleasing. Many candidates made it look very straightforward. Most realised that the non-Algebra books did not have to be subdivided. Some used the longer approach of working out the different combinations of Calculus and Trigonometry books separately, but often still obtained the right answer. A few tried to multiply probabilities, and most of those who did so forgot to multiply their result by ${}^5C_3$.

Question 6 (b)

The teacher now arranges all 10 books in random order on a shelf.

(b) Find the probability that the Calculus books are next to each other and the Trigonometry books are next to each other. **[3]**

Although explanations were rarely as clear as they might have been, there was also a pleasing number of correct answers here. In such questions candidates might find it helpful to refer to "gaps" or "barriers" and not just to put in unexplained symbols on a diagram. As usual, it is rarely a successful strategy to attempt a question such as this by multiplying probabilities; those who did so usually did not realise that the product of five probabilities then had to be multiplied by the number of ways in which they could be arranged, which in this case was ${}^7P_2$.

Question 6 (c)

In this question you must show detailed reasoning.

(c) Find the probability that 2 of the Calculus books are next to each other but the third Calculus book is separated from the other 2 by at least 1 other book. **[4]**

This was a deliberately harder question. There were two different possible approaches. The more usual was to calculate directly the number of ways in which one gap could be occupied by two Calculus books and a different gap by the third Calculus book. They then had to multiply by the number of arrangements of the other 7 books and divide by 10!. The alternative approach was to calculate the probability that all three Calculus books were together (which is routine), and the probability that all three were separate (using gaps), before subtracting these two probabilities from 1. Most of the relatively few successful candidates used the former approach.

Question 7 (a)

7 Each of three students, X, Y and Z, was given an identical pack of 48 cards, of which 12 cards were red and 36 were blue. They were each told to carry out a different experiment, as follows:

Student X: Choose a card from the pack, at random, 20 times altogether, **with** replacement. Record how many times you obtain a red card.

Student Y: Choose a card from the pack, at random, 20 times altogether, **without** replacement. Record how many times you obtain a red card.

Student Z: Choose single cards from the pack at random, **with** replacement, until you obtain the first red card. Record how many cards you have chosen, including the first red card.

(a) Find the probability that student Z has to choose more than 8 cards in order to obtain the first red card. **[3]**

Most candidates identified the relevant distribution as Geo(0.25) and most obtained the correct probability, though some calculated 0.25^8 rather than 0.75^8.

Question 7 (b) (i)

Each student carries out their experiment 30 times.

The frequencies of the results recorded by each student are shown in the following table, but **not** necessarily with the rows in the order X, Y, Z:

	Number recorded	0	1	2	3	4	5	6	7	8	$\geqslant 9$	Observed Mean	Observed Variance
Observed Frequencies	Student 1	0	0	1	3	7	8	6	4	1	0	5.03	1.97
	Student 2	0	8	5	4	2	3	3	2	1	2	4.03	11.57
	Student 3	0	1	2	5	4	6	5	3	4	0	4.97	3.70

(b) In this question you must show detailed reasoning.

Two other students make the following statements about the results. For each of the statements, explain whether you agree with the statement. Do **not** carry out any hypothesis tests, but in each case you should give **two** justifications for your answer.

(i) "The second row is a good match with the expected results for student Z." **[4]**

Ofqual guidelines for problem-solving in mathematics suggest that problem-solving tasks are characterised as having little or no scaffolding, do not explicitly state the mathematical process(es) required for the solution, and allow a variety of different approaches. This question addressed these characteristics.

However, in order to make proper use of the data it was all but essential to identify the relevant distributions (Geo(0.25) for student Z, B(20, 0.25) for student X). Comparison of the respective observed and expected means and variances was then a natural approach, although arguments based on, for example, probabilities could gain credit if sufficient and accurate calculation was shown. Those who tackled either of these two parts by arguing in general qualitative terms were rarely able to produce convincing arguments.

In 7(b) (i), comparison of, for example, the expected mean and variance from Geo(0.25) with 4.03 and 11.57 enabled many candidates to say that they agreed that the second row is a good match for student Z.

Question 7 (b) (ii)

(ii) "The third row is definitely student X's results." **[3]**

In 7 (b) (ii), similar calculations certainly indicate that B(20, 0.25) is a good match for student X, but the statement was that "the third row is <u>definitely</u> student X's results", and candidates had to deny that this was <u>definitely</u> true; there is no question of certainty here.

Exemplar 3

7(b)(ii)	student $X \sim B(\cancel{20}^{20}, 0.25)$
	$E(X) = 20 \times 0.25 = 5$
	$var(X) = 20 \times 0.75 \times 0.25 = 3.75$
	values recorded for $E(X) = 4.97 \approx 5$
	recorded for $var(X) = 3.7 \approx 3.75$
	I agree with the statement as the values expected for
	the mean which is 5 is extremely close to the 4.97 that was
	found.
	I also agree as the variance expected (3.7~~8~~) was extremely close
	to the variance recorded at 3.7 hence the results almost
	perfectly match the expected ~~variance~~ result. for mean and variance

This candidate has correctly identified the relevant distribution for student X as B(20, 0.25), and has then given an excellent comparison of observed and expected mean and variance. However, the conclusion "I agree with the statement" does not score the final mark. The statement was that "The third row is <u>definitely</u> student X's results", and one cannot be 100% definite about this, however well the data match the model.

Supporting you

Post-results services

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.

Keep up-to-date

We send a weekly roundup 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 on the relevant subject page on our website or visit OCR professional development.

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.

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.

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**
- **/ocrexams**
- **/ocrexams**
- **/company/ocr**
- **/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