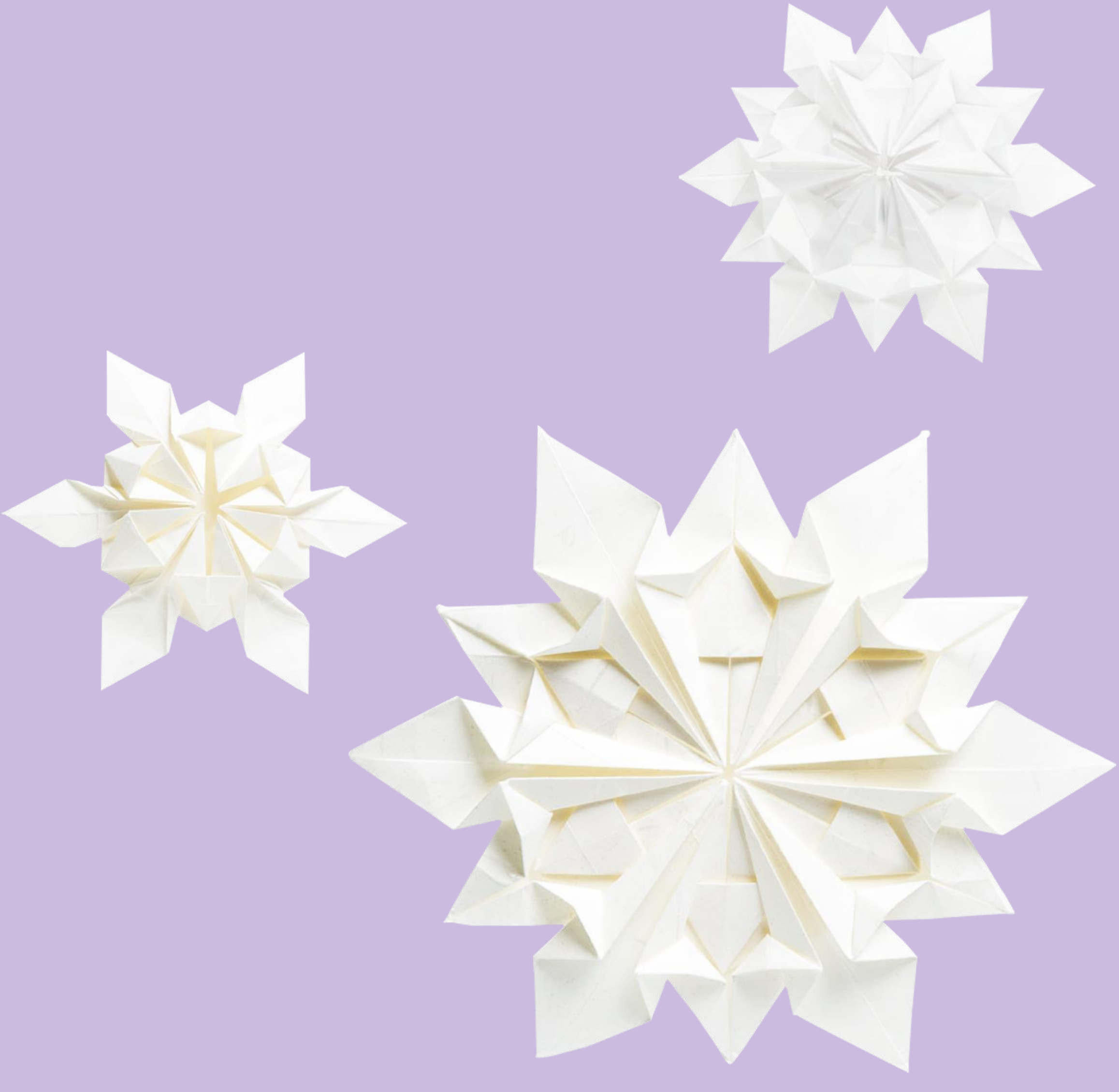
Pearson Edexcel
Level 3 Advanced Subsidiary
GCE in Further Mathematics (8FM0)

Pearson Edexcel
Level 3 Advanced
GCE in Further Mathematics (9FM0)



June 2019 - Further Statistics Exemplar

Student answers with examiner comments

First teaching from September 2017

Contents

About this booklet

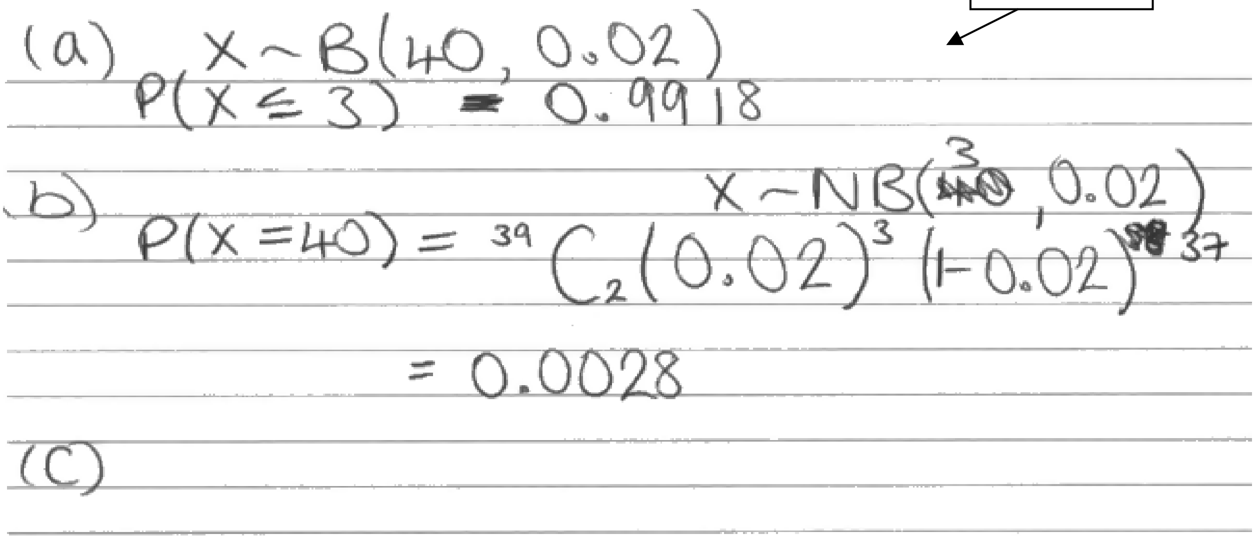
This booklet has been produced to support mathematics teachers delivering the new Pearson Edexcel Level 3 Advanced Subsidiary and Advance Level GCE in Further Mathematics specification (8FM0 & 9FM0). The booklet looks at questions from the AS and A Level Further Mathematics – Further Statistics June 2019 Examination Papers. It shows student responses to questions, and how the examining team follow the mark schemes to demonstrate how the students would be awarded marks on these questions.

How to use this booklet

Our examining team have selected student responses to all questions from the June 2018 Examination Papers. Following each question, you will find the mark scheme for that question and then a range of student responses with accompanying examiner comments on how the mark scheme has been applied and the marks awarded, and on common errors for this sort of question.

Student response

Student Response B



(a) $X \sim B(40, 0.02)$

$P(X \le 3) = 0.9918$

b) $X \sim NB(\cancel{40}\,3, 0.02)$

$P(X=40) = {}^{39}C_2 (0.02)^3 (1-0.02)^{\cancel{38}\,37}$

$= 0.0028$

(c)

3/6

Examiner Comments

In part (a) the correct model is given (M1) but the answer is incorrect (A0)

In part (b) the correct model is stated (M1) and a correct expression is given (M1) but the answer is only given to 2s.f. not 3 and so the last mark was lost (A0)

In part (c) there is no attempt (B0)

Examiner commentary on the student response

Marks awarded for the question or question parts

AS Further Maths – Further Statistics 1 (8FM0 23)

Exemplar Question 1

back to Contents Page

1 A leisure club offers a choice of one of three activities to its 150 members on a Tuesday evening. The manager believes that there may be an association between the choice of activity and the age of the member and collected the following data.

Activity / **Age *a* years**	**Badminton**	**Bowls**	**Snooker**
$a < 20$	9	3	3
$20 \leqslant a < 40$	10	10	14
$40 \leqslant a < 50$	16	15	5
$50 \leqslant a < 60$	15	13	11
$a \geqslant 60$	4	19	3

(*a*) Write down suitable hypotheses for a test of the manager's belief.

(1)

The manager calculated expected frequencies to use in the test.

(*b*) Calculate the expected frequency of members aged 60 or over who choose snooker, used by the manager.

(1)

(*c*) Explain why there are 6 degrees of freedom used in this test.

(2)

The test statistic used to test the manager's belief is 19.583

(*d*) Using a 5% level of significance, complete the test of the manager's belief.

(2)

(Total for Question 1 is 6 marks)

Mean Score 3.7 out of 6

Examiner Comments

Nearly all the students knew the hypotheses were about association or independence and mentioned the context, but a surprising number gave H_0 and H_1 the wrong way around.

Part (b) was answered very well and the critical value in part (d) was usually quoted correctly with these two marks being scored by the vast majority of the students.

The explanation in part (c) was not always answered thoroughly enough. We were looking for students to realise that some of the rows would need amalgamating because some expected frequencies were less than 5. Some calculated all of the expected frequencies but their calculation in part (b) should have steered them towards the top right-hand cell which had an expected frequency of 3.6 and therefore meant that some amalgamation was required. Those who did identify this value usually realised that amalgamating the first two rows would deal with the small expected frequencies and the ensuing 4x3 table would lead to 6 degrees of freedom. Few students calculated the 3.6 (though one mark was available for mentioning the combining or rows leading to the 4x3 table) and some confused rows and columns or just mentioned vaguely combining cells. A common error was to think that amalgamation of cells was required because of the observed frequencies being less than 5 and this sometimes led to combining snooker and bowls or the first and last rows. Others simply thought that the 6 came from a calculation of 8 – 2 presumably confusing with the calculations required for goodness of fit tests.

In the final part, (d), only a small minority did not know how to use the tables and gave the value 1.635 and those who had their hypotheses the correct way round in part (a) were usually able to give an appropriate conclusion either in terms of there being evidence of association between age and choice of activity or, more simply, stating that there was evidence to support the manager’s belief.

Mark Scheme

Qu	Scheme	Marks	AO
1(a)	H_0 : There is no association between age and activity H_1 : There is an association between age and activity	B1 (1)	1.2
(b)	$\frac{26\times 36}{150}$ = 6.24	B1 (1)	1.1b
(c)	Since expected value in *a* < 20 and snooker = 3.6 < 5 we amalgamate two rows	B1	2.4
	Table is now 4×3 so degrees of freedom is $(4-1)\times(3-1)=6$ (*)	B1* (2)	1.1b
(d)	Critical value $\chi_6^{\,2}(5\%)=12.592$	B1	1.1b
	[Significant result]: so there is evidence to support manager's belief	B1ft (2)	2.2b
		(6 marks)	

Notes	
(a)	B1 for both hypotheses in terms of "association" or "independence". Must mention age and activity (or sport). [Use of "relationship" or "link" here is B0 but allow for last B1ft]
(b)	B1 for 6.24
(c)	1st B1 for a reason to get a 4 x 3 table based on amalgamation of rows Must mention *a* < 20 and snooker and see 3.6 and be combining <u>rows</u> (not columns) 2nd B1* for 6 degrees of freedom clearly coming from a 4 x 3 table formed from amalgamation of <u>rows</u>. [8 – 2 = 6 is B0]
(d)	1st B1 for correct critical value (allow 12.6 or 12.59 or awrt 12.592) NB *p*-value = 0.0032839… so allow awrt 0.00328 2nd B1ft for a correct comparison and conclusion (ft their cv) [<u>Independent of hypotheses</u>] e.g. there is an "association" or "relationship" or "link" between age and activity is OK BUT there is a "correlation" between age and activity is B0 Do not accept contradictory contextual statements e.g. "manager's belief supported, there is no association between age and activity"

Examiner Comments

The Assessment Objective addressed in part (c) requires students to give an explanation and many lost the first mark for failing to identify which cell was the cause of the need for pooling of rows.

The final conclusion in part (d) should always be given using the context of the question, so simply stating that the result was "significant" or you "reject H_0" was not sufficient.

Student Response A

(a) H_0: there is an association between the choice of activity and the age of the member.

H_1: there is no association between the choice of activity and the age of the member.

(b). $\frac{\text{row} \times \text{column}}{\text{total}} = \frac{(4+19+3)(3+14+5+11+3)}{150}$

$= \frac{(26)(36)}{150} = 6.24$

(c) (row − 1)(column − 1)

must take 1 from each for constraints.

Must take 2 ~~fo~~ away as ~~we~~ have to work out probability.

(d) 5%. test statistic = 19.583.

$\nu = 6$

$\chi^2_6(5\%) = 12.5915872$

$12.592 < 19.583$ ∴ reject H_0, there is sufficient evidence to show that there is no association between choice of activity and age of member – the managers belief is rejected.

2/6

Examiner Comments

In part (a) the hypotheses are the wrong way around (B0).

In part (b) the calculation is correct here and the answer of 6.24 is seen (B1).

In part (c) they do not seem to realise that the 6 degrees of freedom comes from a 4x3 table and this table comes about after amalgamating two rows (B0B0).

In part (d) the correct critical value is quoted. We provide these values in the tables, but this student has clearly used their calculator to find this value which is fine as long as their answer is at least as accurate as the tables so the first B1 is scored. They have said they reject H_0 and given a conclusion in terms of the context of the manager’s belief, but their conclusion is incorrect because their hypotheses are incorrect. We would always expect a correct conclusion to a hypothesis test and not allow any follow through from incorrect hypotheses (B0).

Student Response B

a) H_0: There is no association between age and choice of sport.
H_1: there is an association between age and choice of sport.

b) Expected 36 × 26 ÷ 150 = 6.24

c) The expected value of a<20 for snooker is less than 5 (3.6), so you combine this column with bowls. Therefore you have (2-1)(5-1) degrees of freedom which is 4. Because you do 1 - the number of cells across and down multiplied together.

d) $\chi^2_4(5\%)$: 9.488. 19.583 > 9.488 therefore there is sufficient evidence at 5% significance level to reject H_0. There is an association between age and sport.

3/6

Examiner Comments

In part (a) the hypotheses are both correct (B1).

In part (b) the correct expected frequency is found (B1).

In part (c) they have identified the cell where the expected frequency is 3.6 and therefore less than 5. But they have suggested combining the columns of snooker and bowls. They go on to show that this would give a 2x5 table and therefore 4 degrees of freedom. The question clearly asked them to explain why there were 6 degrees of freedom and the first mark required identification of the correct cell and mention of combining rows so that mark cannot be given and the calculation of 6 degrees of freedom had to come from a 4x3 table for the second mark (B0B0).

In part (d) they have persisted with using 4 degrees of freedom and so the critical value is incorrect (B0) but their conclusion is correct and in the context of age and choice of activity, so the final mark is scored (B1).

Student Response C

(2)

1) a) H_0 : There is no association between the choice of activity and the age of the member on a tuesday evening

H_1: There is an association between the choice of activity and the age of the member on a tuesday evening

b) $E(i) = \frac{(3 + 14 + 5 + 11 + 3)(4 + 19 + 3)}{150}$

$= 6.24$

c) Since cell 'Snooker $\cap$ $a < 20$' is < 5, has ~~to be co~~ $a < 20$ has to be combined with $20 \leq a < 40$

$\therefore v = (3-1)(4-1)$
$= 6$

d) ~~H_0: there~~ Reject H_0 if $\sum \frac{(O_i - E_i)^2}{E_i} > X^2_6(5\%)$

~~19~~ $X^2_6(5\%) = 12.592$

$19.583 > 12.592 \therefore$ Reject H_0.

There is sufficient evidence to suggest an association between the choice of activity and the age of the member on a tuesday evening at a 5% level of significance.

✓ H_0: there is no association between the choice of activity and the age of the member on a tuesday evening.

H_1: There is an association between the choice of activity and the age of the member on a tuesday evening.

5/6

Examiner Comments

In part (a) the hypotheses are correct (B1).

In part (b) the correct expected frequency is given (B1).

In part (c) they have identified the correct cell and stated that the expected frequency is less than 5 but have not given the value (3.6) of this expected frequency so we cannot be sure that their reason is correct and so the first mark is not awarded. However, they do show that the 6 degrees of freedom clearly comes from a 4x3 table (B0B1).

In part (d) they have a correct critical value and a correct conclusion using the context of age and activity rather than the manager’s belief, but this is fine (B1B1).

back to Contents Page

Exemplar Question 2

2. A spinner used for a game is designed to give scores with the following probabilities

Score	1	2	3	4	6
Probability	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{5}$	$\frac{1}{10}$

The spinner is spun 80 times and the results are as follows

Score	1	2	3	4	6
Frequency	15	4	12	41	8

Test, at the 10% level of significance, whether or not the spinner is giving scores as it is designed to do. Show your working and state your hypotheses clearly.

(7)

(Total for Question 2 is 7 marks)

Mean Score 5.3 out of 7

Examiner Comments

This question was generally answered very well. Some lost marks for the hypotheses: statements such as the “observed frequencies match the expected frequencies” or “a uniform distribution is a suitable model” were given rather than a comment about the spinner’s design such as “the spinner is giving probabilities as it is designed to do”. The expected frequencies were usually found correctly but a number of students lost marks for incorrectly combining the scores for 1 and 2 because the observed frequency for a score of 2 was less than 5; students need to understand that the amalgamation of cells is only required when expected frequencies fall below 5. The calculations and selection of the critical value were usually correct and most realised that the test suggested the spinner was not working as designed.

The final mark here was dependent on all previous marks being scored. This is a fairly common situation in questions involving hypothesis tests. In this question a number of students therefore lost the first and last marks

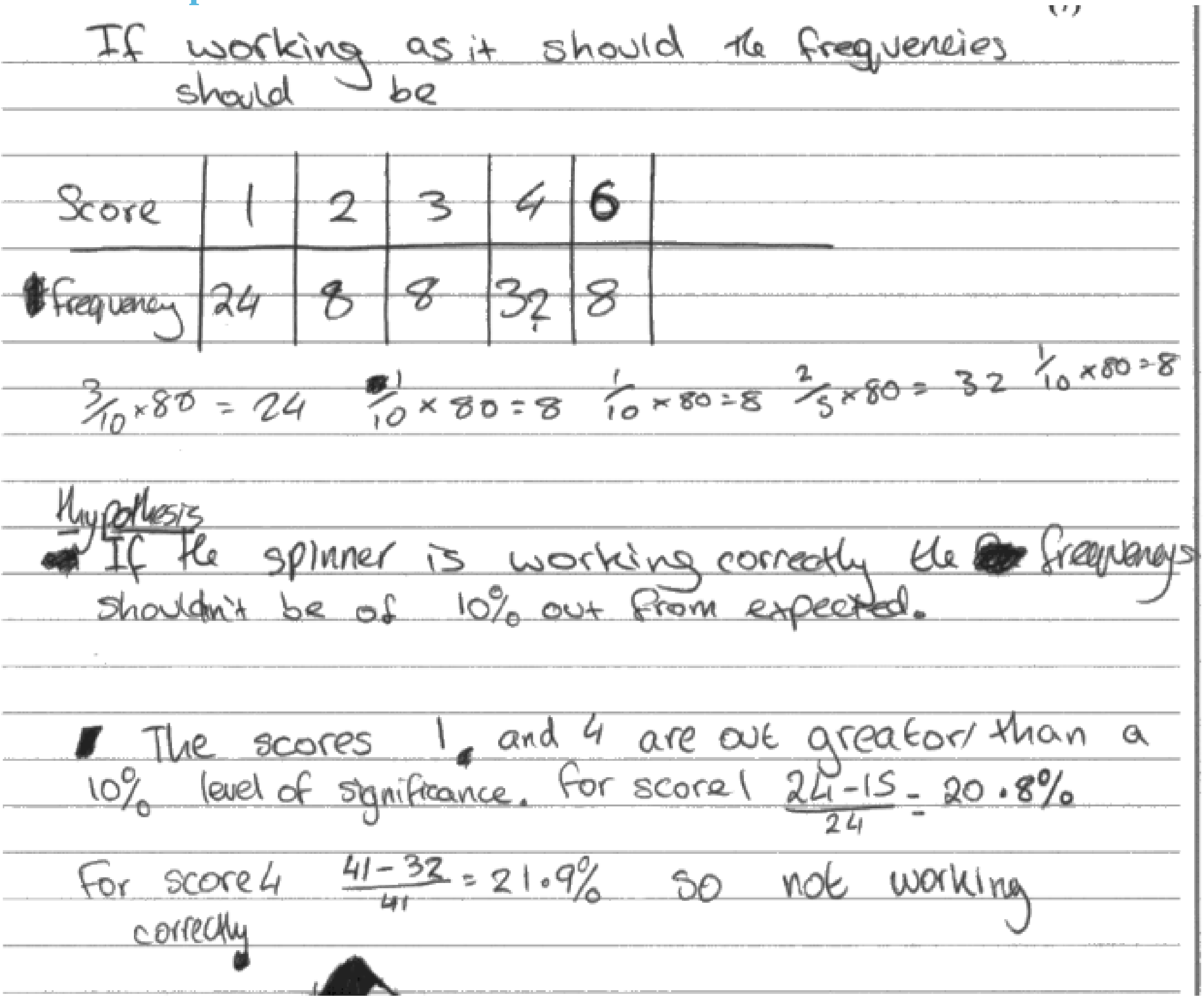
Mark Scheme

<table>
<tr><th>Qu</th><th>Scheme</th><th>Marks</th><th>AO</th></tr>
<tr><td>2.</td><td>H_0: Spinner is working as designed (o.e.)
H_1: Spinner is not working as designed (o.e.)</td><td>B1</td><td>1.2</td></tr>
<tr><td></td><td>
<table>
<tr><td>E_i</td><td>24</td><td>8</td><td>8</td><td>32</td><td>8</td></tr>
<tr><td>O_i</td><td>15</td><td>4</td><td>12</td><td>41</td><td>8</td></tr>
<tr><td>$\frac{(O_i-E_i)^2}{E_i}$</td><td>$\frac{81}{24}$</td><td>$\frac{16}{8}$</td><td>$\frac{16}{8}$</td><td>$\frac{81}{32}$</td><td>0</td></tr>
<tr><td>$\frac{O_i^2}{E_i}$</td><td>$\frac{225}{24}$</td><td>$\frac{16}{8}$</td><td>$\frac{144}{8}$</td><td>$\frac{1681}{32}$</td><td>$\frac{64}{8}$</td></tr>
</table>
</td><td>M1
A1
M1</td><td>3.4
1.1b
1.1b</td></tr>
<tr><td></td><td>$\sum\frac{(O_i-E_i)^2}{E_i}=3.375+2+2+2.53125+0=9.90625$
<u>or</u> $\sum\frac{O_i^2}{E_i}-N$ = 9.375 + 2 + 18 + 52.53125 + 8 – 80 = 9.90625</td><td>A1</td><td>1.1b</td></tr>
<tr><td></td><td>$\nu=5-1=4$ so $\chi_4^2(10\%)$ cv = 7.779 or better</td><td>B1</td><td>3.4</td></tr>
<tr><td></td><td>Result is significant so there is evidence that the spinner is not operating as designed</td><td>A1cso</td><td>3.5a</td></tr>
<tr><td></td><td></td><td>(7 marks)</td><td></td></tr>
<tr><th colspan="4">Notes</th></tr>
<tr><td></td><td colspan="3">1st B1 for both hypotheses given in suitable context
1st M1 for using the model to find at least 2 correct expected frequencies
1st A1 for all correct E_i
2nd M1 for attempt to find test statistic (at least two correct expressions, fractions or decimals)
2nd A1 for a correct test statistic (awrt 9.91) [accept $\frac{317}{32}$]
2nd B1 for correct critical value (allow 7.78)
NB p- value = 0.042036…so allow awrt 0.042
3rd A1cso dep on <u>all</u> previous marks for a correct conclusion in context
(can be in terms of model or spinner's design) Must mention spinner and scores <u>or</u> design
Accept "spinner is not accurate"</td></tr>
</table>

Examiner Comments

The final mark here was dependent on all previous marks being scored. This is a fairly common situation in questions involving hypothesis tests. In this question a number of students therefore lost the first and last marks.

Student Response A



If working as it should the frequencies should be

Score	1	2	3	4	6
Frequency	24	8	8	32	8

$\frac{3}{10} \times 80 = 24$ $\frac{1}{10} \times 80 = 8$ $\frac{1}{10} \times 80 = 8$ $\frac{2}{5} \times 80 = 32$ $\frac{1}{10} \times 80 = 8$

Hypothesis

If the spinner is working correctly the frequencys shouldn't be of 10% out from expected.

The scores 1, and 4 are out greator than a 10% level of significance. For score 1 $\frac{24-15}{24} = 20.8\%$

For score 4 $\frac{41-32}{41} = 21.9\%$ so not working correctly

2/7

Examiner Comments

There was no clear statement of the hypotheses (B0) but the expected frequencies were all correct (M1A1). This student hasn't attempted to carry out a Chi-squared goodness of fit test but has simply shown that the differences between some of the expected and observed frequencies are greater than 10% so no further marks were scored (M0A0B0A0).

Student Response B

(i)

H_0 : The spinner is unsuitable

H_1 : The spinner is suitable

Expected:

S	1	2	3	4	~~5~~ 6
E	24	8	8	16	8

S	1	2	3	4	6
$\frac{(O-E)^2}{E}$	3.375	2	2	39.0625	0

$(S-1) = 4$ DoF

$$\Sigma \chi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) = \frac{743}{16} = 46.4375$$

$$\chi^2_4(10\%) = 7.779$$

∴ not in CR ∴

$46.4375 > 7.779$ ∴ don't rej H_0 ∴ there is no evidence to suggest that the spinner is suitable.

3/7

Examiner Comments

The hypotheses are not given in terms of the spinner’s design. By assuming the null hypothesis is true we should be able to calculate the expected frequencies: just saying the spinner is “suitable” does not give us sufficient information to do this (B0). The expected frequencies for scores of 1, 2, 3 and 6 are correct but the value given for 4 is not (M1A0) and the values for the test statistic for these values are also correct (M1A0). The critical value is correct (B1) and because other marks have been lost the final mark cannot be given (A0). Their interpretation of their critical value and their tests statistic is incorrect as well as a test statistic value greater than a critical value should lead to rejection of the null hypothesis.

Student Response C

(7)

H_0: The spinner is giving probabilities of scores it should.

H_1: The spinner is not giving the right probabilities of scores. Not working as designed to.

	1	2	3	4	6
Observed:	15	4	12	41	8
expected:	24	8	8	32	8
contribution	3.375	2	2	2.53125	0

$\chi^2 = 9.90625$

Expected = probability × 80

$$\text{Contribution} = \frac{(o-e)^2}{e}$$

degrees of freedom $= 5-1 = 4$

$\chi^2_4(0.1) = 7.779$

$9.40625 > 7.779$

~~Reject~~ Not sufficient evidence to reject Ho.

Accept that it is working how it was designed.

6/7

Examiner Comments

The hypotheses are correct (B1). The expected frequencies are correct (M1A1) and the calculation of the test statistic is correct too (M1A1). The critical value is correct (B1) but they have said that there is not sufficient evidence to reject H_0 which is incorrect (A0).

Exemplar Question 3

back to Contents Page

3. Andreia’s secretary makes random errors in his work at an average rate of 1.7 errors every 100 words.

(*a*) Find the probability that the secretary makes fewer than 2 errors in the next 100-word piece of work.

(2)

Andreia asks the secretary to produce a 250-word article for a magazine.

(*b*) Find the probability that there are exactly 5 errors in this article.

(2)

Andreia offers the secretary a choice of one of two bonus schemes, based on a random sample of 40 pieces of work each consisting of 100 words.

In scheme **A** the secretary will receive the bonus if more than 10 of the 40 pieces of work contain no errors.

In scheme **B** the bonus is awarded if the total number of errors in all 40 pieces of work is fewer than 56

(*c*) Showing your calculations clearly, explain which bonus scheme you would advise the secretary to choose.

(5)

Following the bonus scheme, Andreia randomly selects a single 500-word piece of work from the secretary to test if there is any evidence that the secretary’s rate of errors has decreased.

(*d*) Stating your hypotheses clearly and using a 5% level of significance, find the critical region for this test.

(4)

(Total for Question 3 is 13 marks)

Mean Score 9.3 out of 13

Examiner Comments

A few students did not realise the significance of the terms “random” and “average rate” and used binomial rather than Poisson models throughout the question but for most students this question provided a good source of marks. In part (a) the most frequent error was a mis-interpretation of the phrase “fewer than 2” with many simply finding the probability of less than or equal to 2.

Responses to part (b) were often better and most found the new Poisson distribution and used it correctly.

Part (c) was the most challenging part of the question with many unable to work out the probability connected with scheme **A**. Most realised there was something to do with a Po(1.7) model but far fewer realising that finding P($X = 0$), where X~Po(1.7), was giving them a parameter in a binomial model. Those who realised this often failed to use sufficient accuracy to achieve their final probability correct to 3 significant figures and Y~ B(40, 0.183) was commonly used. The next problem was interpreting “more than 10” correctly and many failed to use 1 – P($Y \leqslant 10$). There was often more success with the probability associated with scheme **B** and those who had correct probabilities were usually able to correctly recommend scheme **A**.

There were many correct answers to part (d) with hypotheses usually given correctly in terms of λ or μ but the common errors were to leave the final answer as P($E \leqslant 3$) rather than $E \leqslant 3$ or confusing a critical region with a critical value and having a final incorrect answer of $E = 3$.

Mark Scheme

Qu	Scheme	Marks	AO
3(a)	[X = number of errors in 100-word piece] $X \sim \text{Po}(1.7)$	M1	3.3
	$\text{P}(X < 2) = \text{P}(X \leqslant 1) = 0.49324...$ awrt **<u>0.493</u>**	A1	1.1b
		(2)	
(b)	[R = number of errors in the article] $R \sim \text{Po}(4.25)$	M1	3.3
	$\text{P}(R = 5) = 0.16482\ldots$ awrt **<u>0.165</u>**	A1	1.1b
		(2)	
(c)	Scheme **A**: Let $A \sim \text{B}(40,\ \text{e}^{-1.7})$ or $\text{B}(40, 0.18268\ldots)$	M1	3.3
	$\text{P}(A > 10) = 1 - \text{P}(A \leqslant 10)$	M1	1.1b
	$= 0.0995591\ldots$ awrt **<u>0.0996</u>**	A1	1.1b
	Scheme **B** : Let $B \sim \text{Po}(40 \times 1.7)$ or $\text{Po}(68)$	M1	3.3
	$\text{P}(B < 56) = \text{P}(B \leqslant 55) = 0.061133\ldots$		
	So choose scheme **A** (since the probability of a bonus is greater)	A1	2.4
		(5)	
(d)	$\text{H}_0 : \lambda = 1.7$ (or $\mu = 8.5$) $\text{H}_1 : \lambda < 1.7$ (or $\mu < 8.5$)	B1	2.5
	[E = no. of errors in the piece of work] $E \sim \text{Po}(8.5)$	M1	3.3
	$\text{P}(E \leqslant 3) = 0.0301$ <u>or</u> $\text{P}(E \leqslant 4) = 0.0744$	A1	1.1b
	So critical region is $\boldsymbol{E \leqslant 3}$	A1	2.2a
		(4)	
		(13 marks)	
	Notes		
(a)	M1 for selecting the correct Poisson distribution		
	A1 for awrt 0.493		
(b)	M1 for selecting the correct Poisson distribution		
	A1 for awrt 0.165		
(c)	1st M1 for choosing a correct model for scheme **A** i.e. B(40, P(X = 0)), where $X \sim \text{Po}(1.7)$		
	Allow use of awrt 0.183 for P(X = 0) … 0.183 gives answer awrt 0.101		
	Condone B(0.183, 40) (o.e.) if it leads to a prob rounding to range (0.09~0.1) otherwise M0		
	2nd M1 for $1 - \text{P}(A \leqslant 10)$		
	1st A1 for awrt 0.0996 [NB use of 0.183 will give awrt 0.101 and scores M1M1A0]		
	3rd M1 for selecting a correct Poisson model for scheme **B** i.e. $\text{Po}(40 \times 1.7)$ or better		
	2nd A1 for a correct conclusion based on comparing two probs: awrt 0.1 vs 0.061 or better		
	So can allow 0.1 > 0.061 leading to choosing **A** [Probably scores M1M1A0M1A1]		
NB	[Normal approx.(not on spec) leading to 0.06477… might score 3rd M1 if Po(68) seen but 2nd A0]		
(d)	B1 for both hypotheses in terms of λ or μ (can be interchanged)		
	M1 for selecting Po(8.5) (sight of or use of e.g. may be implied by 1st A1)		
	1st A1 for some evidence of correct use of Po(8.5) i.e. either of these probs (2dp or better)		
	May be implied by a correct critical region		
	2nd A1 for a correct critical region. Allow $E < 4$ and allow any letter for E.		
	<u>Two</u> different regions (e.g. from 2 tail test) is 2nd A0		
SC	**Use of binomial throughout:** (with hypotheses $\text{H}_0 : p = 0.017$ and $\text{H}_1 : p < 0.017$ in (d))		
	Scores 0 in (a) 0 in (b) possibly just 2nd M1 in (c) But allow all 4 marks in (d): B1 hypotheses, M1 for $Y \sim \text{B}(500, 0.017)$, 1st A1 for $\text{P}(Y \leqslant 3) = 0.02913\ldots$<u>or</u> $\text{P}(Y \leqslant 4) = 0.07266\ldots$2nd A1 $Y \leqslant 3$		
	Allow probs to be to 2dp or better so 0.03 and 0.07 as in main scheme.		

Examiner Comments

We allow hypotheses to be given in terms of λ (the rate of the Poisson distribution) or μ (the mean of the distribution).

Student Response A

$X \sim Po(1.7)$ every 100 words

a) $P(X < 2) = P(X \leq 1) = \underline{\underline{0.49}}$

b) 250 words $X \sim Po(4.25)$

$P(X = 5) = e^{-4.25}\dfrac{4.25^5}{5!} = \underline{\underline{0.16}}$

c) Scheme A

$X \sim Po(1.7)$

$P(X = 0)$ is ~~[illegible]~~ ~~[illegible]~~ more than 10

$P(X \leq 0) = 0.18$

~~[illegible]~~ $0.18^{11} = 6.4 \times 10^{-9}$

~~$0.18^{10} = 3.57 \times 10^{-8}$~~

Scheme B

~~P(X~~ $X \sim Po(68)$

$P(X < 56) = P(X \leq 55) = 0.061$

$6.4 \times 10^{-9} < 0.061$

∴ Use scheme B as there is a higher probability of achieving this goal

d) $H_0 : \lambda = 8.5$

$H_1 : \lambda < 8.5$

5% sig level

$P(X \leq 4) = 0.074 > 0.05$

$P(X \leq 3) = 0.03 < 0.05$

critical region $\underline{X \leq 4}$

5/13

Examiner Comments

In part (a) the correct Poisson model is stated (M1) but the answer is not given to 3s.f. (A0).

In part (b) again the correct model is used but the answer is not given to 3s.f. despite the clear instruction to this effect on the front of the question paper (M1A0).

In part (c) for scheme **A** they have used Po(1.7) to find P(X = 0) but have not realised that the distribution they then need is a binomial (M0) and there is no clear statement of the correct probability (M0A0). For scheme **B** they do use the correct distribution (M1) but the final mark requires a comparison of two correct probabilities (A0).

In part (d) we allow $\lambda = 8.5$ and $\lambda < 8.5$ in the hypotheses (B1) and there is at least 1 correct probability (2dp accuracy is fine here as they only need to be comparing to 0.05) (M1A1) but the final critical region is incorrect (A0).

Student Response B

(2)

a) Ho: ~~Age~~ There is an association between the choice of activity and the age of the member
H₁: There is no association between age and activity

B) 4 members

C) ~~[illegible]~~ There are 5 rows in the table and then you ~~[illegible]~~ add 1 to calculate the degrees of freedom because you use the data once, therefore ~~[illegible]~~ it is 6.

D) 5% sig = 0.05

$X \sim Po(4)$

7/13

Examiner Comments

In part (a) this is correct (M1A1).

In part (b) this is correct (M1A1).

In part (c) the correct binomial model for scheme **A** is not identified (M0M0A0) but the correct model for scheme **B** is used (M1) but there is no comparison of two correct probabilities (A0).

In part (d) the hypotheses are correct (B1) and a correct Poisson model is stated (M1) but there are no correct probability statements and the critical region is not correct (A0A0).

Student Response C

$X \sim Po(1.7)$ per 100 words

(a) $P(X < 2) = P(X \leq 1) = 0.493$ (3sf)

(b) $X \sim Po(1.7 \times 2.5)$

$X \sim Po(4.25)$

$$P(X=5) = e^{-4.25} \times \frac{4.25^5}{5!} = 0.165 \ (3sf)$$

P(no errors in 100 words) → $P(X=0) = 0.183$ (3sf)

$X \sim Po(1.7)$

In scheme A $\quad X \sim B(40, 0.183)$

$P(X \geq 10) = P(X \geq 11) = 1 - P(X \leq 10)$

$= 1 - 0.899 = 0.101$ (3sf)

(P of getting bonus in scheme 1 $= 0.101$)

In scheme B

Total number of errors is < 56 in 40 pieces of work

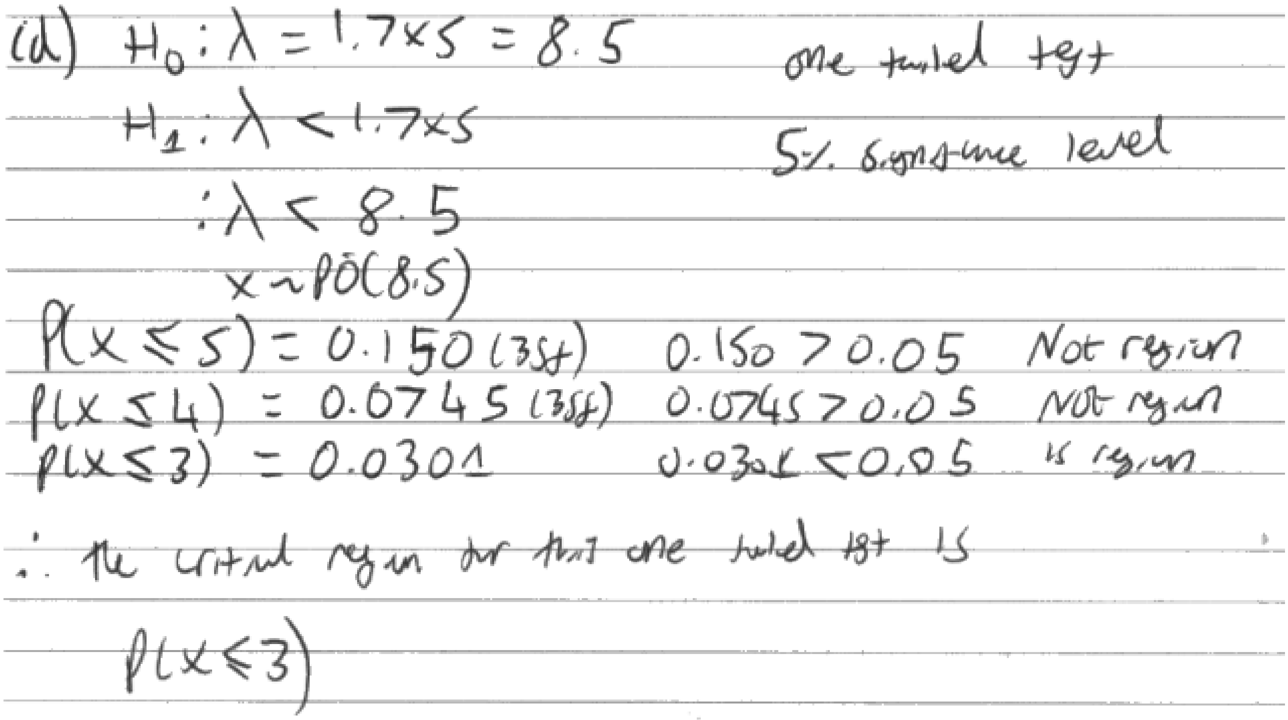
40 pieces $= 40 \times 100$ words $= 4000$ words

$X \sim Po(1.7)$ per 100 $\quad X \sim Po(1.7 \times 40)$ per 4000 words

$X \sim Po(68)$

$P(X < 56) = P(X \leq 55) = 0.0611$ (3sf)

So she has a better probability of getting the bonus in scheme A as $0.101 > 0.0611$ so I would advise the secretary to choose scheme A.



(d) $H_0 : \lambda = 1.7 \times 5 = 8.5$ one tailed test

$H_1 : \lambda < 1.7 \times 5$ 5% significance level

$\therefore \lambda < 8.5$

$X \sim Po(8.5)$

$P(X \leq 5) = 0.150$ (3sf) $0.150 > 0.05$ Not region

$P(X \leq 4) = 0.0745$ (3sf) $0.0745 > 0.05$ Not region

$P(X \leq 3) = 0.0301$ $0.0301 < 0.05$ is region

$\therefore$ the critical region for that one tailed test is

$P(X \leq 3)$

11/13

Examiner Comments

In part (a) this is correct (M1A1).

In part (b) this is correct (M1A1).

In part (c) for scheme **A** they have identified and used an appropriate binomial model (M1M1) but the value for P(X = 0) they have used as the parameter was not sufficiently accurate to enable them to get a final probability correct to 3s.f. (A0). The correct probability for scheme **B** is found (M1) and a correct comparison and conclusion is made (A1) [We allow comparison of a probability of 0.1 with 0.061 for this mark].

In part (d) the hypotheses are correct (B1) and the correct model is used to obtain some correct probabilities (M1A1) but the critical region is written inside a probability statement, so the final mark is lost (A0).

Exemplar Question 4

back to Contents Page

4. The discrete random variable X has probability distribution

x	-3	-1	1	2	4
$\mathrm{P}(X = x)$	q	$\frac{7}{30}$	$\frac{7}{30}$	q	r

where q and r are probabilities.

(*a*) Write down, in terms of q, $\mathrm{P}(X \leqslant 0)$

(1)

(*b*) Show that $\mathrm{E}(X^2) = \frac{7}{15} + 13q + 16r$

(2)

Given that $\mathrm{E}(X^3) = \mathrm{E}(X^2) + \mathrm{E}(6X)$

(*c*) find the value of q and the value of r

(7)

(*d*) Hence find $\mathrm{P}(X^3 > X^2 + 6X)$

(4)

(Total for Question 4 is 14 marks)

Mean Score 10.4 out of 14

Examiner Comments

Parts (a) and (b) were answered very well but a few students miscopied their $16r$ as $6r$ and proceeded to use this incorrect value in part (c).

Most could make a start with part (c) by finding $\mathrm{E}(X)$ and hence $6\mathrm{E}(X)$ and $\mathrm{E}(X^2) + \mathrm{E}(6X)$. Many also managed to find an expression for $\mathrm{E}(X^3)$ though sometimes there were sign errors on the negative terms. Finding a second equation in q and r seemed to evade a few but most used the sum of the probabilities to form a second equation and were able to solve their two equations which, if no previous errors had been made, usually led to the correct answers.

There were a few blank responses to part (d) and a few who tried comparing expected values and then gave up, but a good number made a start on part (d) and quite a few of them, with the help of the final A1 follow through mark, achieved all 4 marks. Some started by rearranging the inequality and factorising the cubic but often they stopped here seemingly unable to draw a simple sketch and solve the cubic inequality. The most common approach though was to form a table with the 10 values for X^3 and $X^2 + 6X$ and by comparing these values to identify $X = -1$ and 4 as the required cases and then write down the appropriate probability.

Mark Scheme

Qu	Scheme	Marks	AO
4(a)	$\underline{q+\frac{7}{30}}$	B1	1.1b
		(1)	
(b)	$\mathrm{E}\left(X^2\right)=(-3)^2\times q \;+\; (-1)^2\times\frac{7}{30} \;+\; 1^2\times\frac{7}{30} \;+\; 2^2\times q \;+\; 4^2\times r$	M1	1.1b
	$= \underline{\frac{7}{15}+13q+16r}$ **(*)**	A1*cso	1.1b
		(2)	
(c)	$\mathrm{E}(X)=-3q \;+\; -\frac{7}{30} \;+\; \frac{7}{30} \;+\; 2q \;+\; 4r \quad \{=4r-q\}$	M1	3.1a
	$\mathrm{E}\left(X^2+6X\right)=\frac{7}{15}+7q+40r$	A1	1.1b
	$\mathrm{E}\left(X^3\right)=(-3)^3\times q \;+\; (-1)^3\times\frac{7}{30} \;+\; 1^3\times\frac{7}{30} \;+\; 2^3\times q \;+\; 4^3\times r$	M1	3.4
	$= 64r-19q$	A1	1.1b
	Sum of probabilities = 1 gives: $2q+r=\frac{16}{30}$ (o.e.)	M1	1.1b
	Solve: $24r-26q=\frac{7}{15}$ and $r+2q=\frac{8}{15}$ e.g. $37r=\frac{111}{15}$	dM1	1.1b
	So $\underline{r=\frac{1}{5}}$ and $\underline{q=\frac{1}{6}}$	A1	1.1b
		(7)	
(d)	$X^3>X^2+6X \quad \Rightarrow X(X-3)(X+2)>0$	M1	2.1
	Use of sketch or table to see: $-2<X<0$ or $X>3$	A1	1.1b
	So $\mathrm{P}(X^3>X^2+6X)=\mathrm{P}(X=-1 \text{ or } 4)$	M1	2.2a
	$= \frac{7}{30}+"r"= \underline{\frac{13}{30}}$	A1ft	1.1b
		(4)	

ALT

X	-3	-1	1	2	4
X^3	-27	-1	1	8	64
X^2+6X	-9	-5	7	16	40

(14 marks)

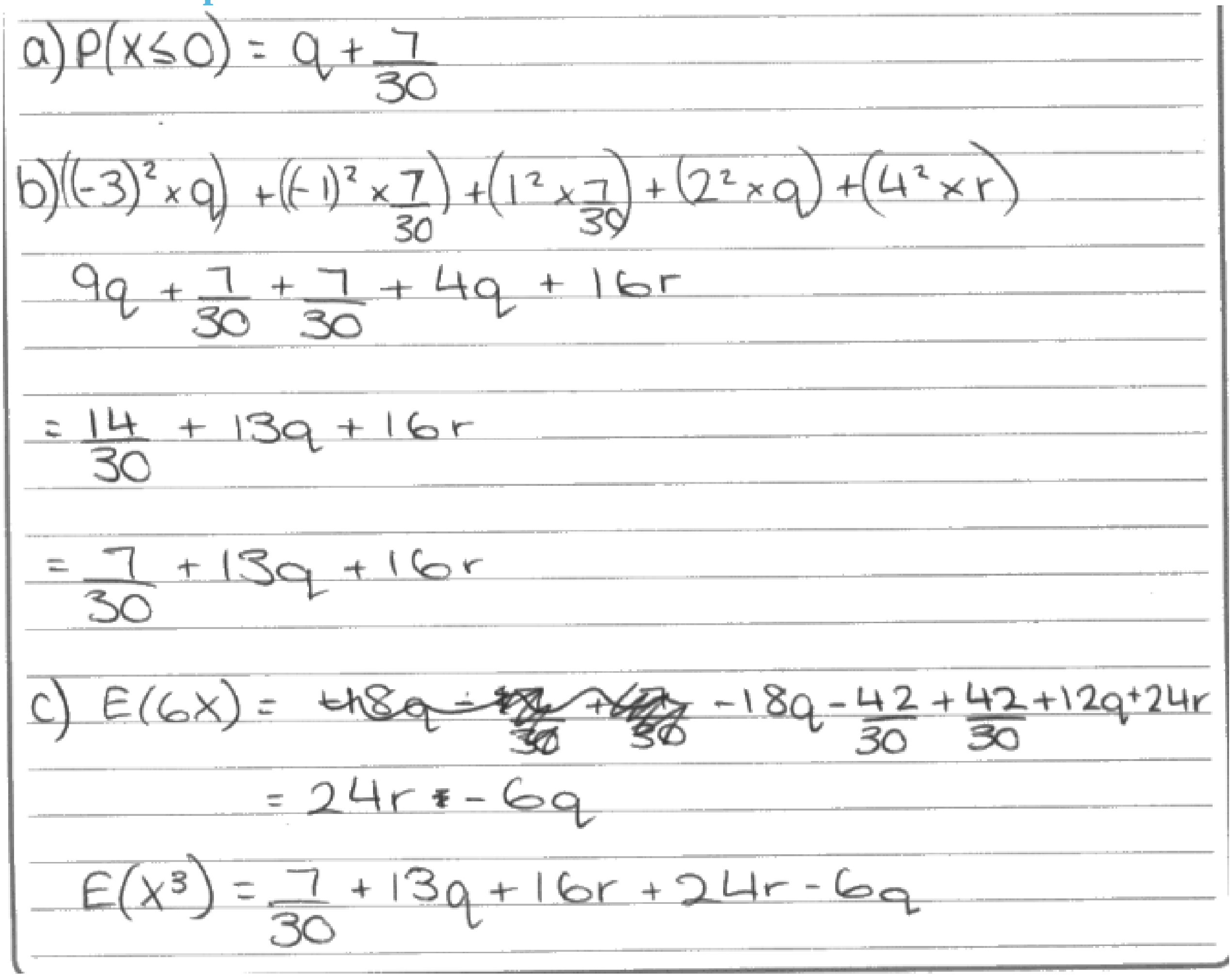
Notes	
(b)	M1 for at least 3 correct terms of the expression for E(X^2) A1*cso evidence of M1 scored with no incorrect working seen leading to correct answer **(*)** Allow $-3^2 \times q + -1^2 \times \frac{7}{30}$ etc if followed by 9q + …but not if simply followed by given answer
(c)	1st M1 for realising the need to find E(X) – a correct attempt with at least 3 correct terms 1st A1 for the correct expression (needn't be simplified at this stage) 2nd M1 for a correct attempt at E(X^3) with at least 3 correct terms seen Treat no $\frac{7}{30}$ terms as one correct term 2nd A1 for 64r – 19q (must be simplified) or for $24r - 26q = \frac{7}{15}$ 3rd M1 for using sum of probabilities = 1 to form an equation in q and r (needn't be simplified) Must be correct or clearly state that Σprobs = 1 being attempted with only one slip 4th dM1 for solving their 2 linear equations in q and r (dep on 3rd M1 and 1st or 2nd M1) Must see correct method to reduce to a linear equation in one variable 3rd A1 for $r = \frac{1}{5}$ and $q = \frac{1}{6}$ or any exact equivalents (dep on 2 correct equations seen)
(d)	1st M1 for 1st stage towards solving the inequality (factorising the cubic) 1st A1 for solving the inequality 2nd M1 for identifying the values of X required i.e. – 1 and 4 2nd A1ft for $\frac{13}{20}$ or exact equivalent e.g. $0.4\dot{3}$ (Allow ft of "their r" + $\frac{7}{30}$)
ALT	**Table** 1st M1 for at least 4 correct values for X^3 and $X^2 + 6X$ (must be labelled) 1st A1 for all 10 correct values. [NB Can score M1A0M1A1ft in (d)]

Examiner Comments

Part (b) was a "show that" question and most students did show clearly how to find E(X^2) showing at least 3 of the products as required by the mark scheme.

The final part gave opportunity for students to use some of the work from pure mathematics to solve the inequality, but many chose instead to construct an exhaustive list of possibilities which was, of course, in this case quite acceptable.

Student Response A



a) $P(X \leq 0) = q + \frac{7}{30}$

b) $((-3)^2 \times q) + ((-1)^2 \times \frac{7}{30}) + (1^2 \times \frac{7}{30}) + (2^2 \times q) + (4^2 \times r)$

$9q + \frac{7}{30} + \frac{7}{30} + 4q + 16r$

$= \frac{14}{30} + 13q + 16r$

$= \frac{7}{30} + 13q + 16r$

c) $E(6X) = \cancel{+18q - \frac{\text{[illegible]}}{30} + \frac{\text{[illegible]}}{30}} - 18q - \frac{42}{30} + \frac{42}{30} + 12q + 24r$

$= 24r \cancel{+} - 6q$

$E(X^3) = \frac{7}{30} + 13q + 16r + 24r - 6q$

$= \frac{7}{30} + 7q + 30r$

$E(X^3) = -27q + 8q + 64r$

~~$-27q$~~

$-19q + 64r = \frac{7}{30} + 7q + 30r$

$-26q + 34r = \frac{7}{30}$

$2q + r = 1 - \frac{7}{15}$

$2q + r = \frac{8}{15}$

$r = \frac{8}{15} - 2q$

$-26q + 34\left(\frac{8}{15} - 2q\right) = \frac{7}{30}$

$-26q + \frac{272}{15} - 68q = \frac{7}{30}$

$-94q = -17.9$

$q =$

7/14

Examiner Comments

In part (a) the correct answer is given (B1).

In part (b) a correct expression for E(X^2) is seen (M1) but it is not simplified correctly to the required form (A0). We require careful and accurate work to reach the printed answer and the given answer with 7/15 is never quite reached.

In part (c) they find E(6X) correctly but then use their incorrect expression for E(X^2) so that their final answer for E(X^2 + 6X) is incorrect (M1A0). There is a correct simplified expression for E(X^3) (M1A1) and a first equation for q and r is formed. A second equation is found using the sum of probabilities (M1) and there is clear evidence of solving these two equations (M1) but the first equation is incorrect and so the answers for q and r would be incorrect and their solution stops before these values are found (A0).
N.B. We award the mark for solving their equations because we have seen clear algebraic steps to achieve this: had this simply been done on a calculator with no evidence of method shown we would not have awarded this mark.

There is no attempt at part (d).

Student Response B

4a) $\frac{7}{30} + q$

b)

x	-3	-1	1	2	4
x^2	9	1	1	4	16
$P(X=x)$	q	$\frac{7}{30}$	$\frac{7}{30}$	q	r

$9q + \frac{14}{30} + 4q + 16r$

$= \frac{7}{15} + 13q + 16r$

e)

~~x~~	~~-3~~	~~-1~~	~~1~~	~~2~~	~~4~~
~~$6x$~~	~~-18~~	~~-6~~	~~6~~	~~12~~	~~24~~
~~x^2~~	~~9~~	~~1~~	~~1~~	~~4~~	~~16~~
~~$P(X=x)$~~	~~q~~	~~$\frac{7}{30}$~~	~~$\frac{7}{30}$~~	~~q~~	~~r~~
~~$6x^3$~~	~~-27~~	~~1~~	~~1~~	~~8~~	~~64~~

x	-3	-1	1	2	4
x^3	-27	-1	1	8	64
$6x$	-18	-6	6	12	24
$P(X=x)$	q	$\frac{7}{30}$	$\frac{7}{30}$	q	r

$E(6X) = -18q - \frac{7}{5} + \frac{7}{5} + 12q + 24r$

$= 24r - 6q$

$\frac{7}{15} + 13q + 16r + 24r - 6q$

$E(X^2+6X) = \frac{7}{15} + 7q + 40r$

$E(X^3) = -27q - \frac{7}{30} + \frac{7}{30} + 8q + 64r$

$= -19q + 64r$

~~$-19q + 64r = \frac{7}{15} + 7q + 4r$~~

~~$-12q + 60r = \frac{7}{15}$~~

$\frac{7}{15} + 7q + 40r = -19q + 64r$

Question 4 continued

d) 14q

$\boxed{r = \frac{1}{5}}$

x	-3	-1	1	2	4
x^2+6x	-9	-5	7	16	40
x^3	-27	-1	1	8	64
$P(X=x)$	q	$\frac{7}{30}$	$\frac{7}{30}$	q	r

$\frac{7}{30} + r$

$= \frac{7}{30} + \frac{1}{5}$

$= \frac{13}{30}$

11/14

Examiner Comments

In part (a) this was correct (B1).

In part (b) the expression for E(X^2) is seen (M1) and it is simplified correctly to the given answer (A1).

In part (c) a correct attempt at E(6X) is seen (M1) and this is added to the answer from part (b) to obtain an expression for E(X^2 +6X) (A1). There is also a correct expression for E(X^3) (M1A1) and they combine these to form a first equation in r and q. There is no attempt to use the sum of the probabilities to form the second equation (M0) and so no attempt to solve two simultaneous equations in r and q (M0A0). Although a correct value for r is stated on the next page, we required sight of both values and 2 correct equations to give the mark in (c).

In part (d) there is a clear table showing the values of X^3 and X^2 + 6X for each of the possible X values and a correct probability is written down following this. Whilst a little more explanation would have been desirable, we can clearly see where the probability is coming from and all the marks are awarded for part (d).

Student Response C

④ a) $q + \frac{7}{30}$

b) ~~$E(x) = 3q - \frac{7}{30} + \frac{7}{30} + 2q + 4r$~~

$E(x^2) = 9q + \frac{7}{30} + \frac{7}{30} + 4q + 16r$

$= [illegible] \cdot \frac{7}{15} + 13q + 16r$

c) $E(x^3) = E(x^2) + E(6x)$

$-27q - \frac{7}{30} + \frac{7}{30} + 8q + 64r = \frac{7}{15} + 13q + 16r - 18q$

$- \frac{7}{5} + \frac{7}{5} + 12q + 24r$

~~$6rr - 19$~~ $64r - 19q = \frac{7}{15} + 7q + 40r$

① $24r - 26q = \frac{7}{15}$

$\Sigma P(X = x) = 1$

$\therefore \; 2q + r + \frac{7}{15} = 1$

② $2q + r = \frac{8}{15}$

~~①x = ① $24r - 26q = \frac{7}{15}$~~

~~$-$ ② ×12 [illegible]~~

① + ② ×13 : $24r - 26q = \frac{7}{15}$

$+ \; 13r + 26q = \frac{104}{15}$

$37r = \frac{111}{15}$

$37r = \frac{37}{5}$

$r = \frac{1}{5}$

$q = \frac{\frac{8}{15} - \frac{1}{5}}{2}$

$$\boxed{\begin{aligned} q &= \frac{1}{6} \\ r &= \frac{1}{5} \end{aligned}}$$

d) ~~$P(x^3 > x^2 + 6x) =$~~

x^3	-27	-1	1	8	64
$P(X=x)$	$\frac{1}{6}$	$\frac{7}{30}$	$\frac{7}{30}$	$\frac{1}{6}$	$\frac{1}{5}$

x^2+6x	-9	-5	7	16	28
$P(X=x)$	$\frac{1}{6}$	$\frac{7}{30}$	$\frac{7}{30}$	$\frac{1}{6}$	$\frac{1}{5}$

$\therefore$ ~~$P(x^3$~~ $\therefore P(x^3 > x^2 + 6) = \frac{7}{30} + \frac{1}{5} = \frac{13}{30}$

$P(x^3 > x^2 + 6x) = \frac{13}{30}$

13/14

Examiner Comments

In part (a) the correct answer is given (B1).

In part (b) there is a correct expression for E(X^2) (M1) and it is simplified to the correct answer (A1).

In part (c) the LHS of their equation represents the calculation of E(X^3) (M1) and this is simplified correctly on the next line (A1). The RHS is their answer from (b) + the expression for E($6X$) which is correct (M1A1). On the next page we see clear use of the sum of probabilities = 1 being used (M1) and then a clear demonstration of solving their 2 simultaneous equation (M1) leading to the correct values for r and q (A1).

In part (d) they again use a table of values for X^3 and $X^2 + 6X$ but there is an error in the evaluation when $X = 4$ so an accuracy mark is lost here but it does not affect their method and the final answer is still correct. (M1A0M1A1).

AS Further Mathematics – Further Statistics 2 (8FM0 24)

Exemplar Question 1

back to Contents Page

1. Bara is investigating whether or not the two judges of a skating competition are in agreement. The two judges gave a score to each of the 8 skaters in the competition as shown in the table below.

	Skater							
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Judge 1	71	70	72	62	63	61	57	53
Judge 2	73	71	67	64	62	56	52	53

Bara decided to calculate Spearman's rank correlation coefficient for these data.

(*a*) Calculate Spearman's rank correlation coefficient between the ranks of the two judges.

(4)

(*b*) Test, at the 1% level of significance, whether or not the two judges are in agreement.

(4)

Judge 1 accidentally swapped the scores for skaters *D* and *E*. The score for skater *D* should be 63 and the score for skater *E* should be 62

(*c*) Without carrying out any further calculations, explain how Spearman's rank correlation coefficient will change. Give a reason for your answer.

(2)

(Total for Question 1 is 10 marks)

Mean Score 6.9 out of 10

Examiner Comments

This question is assessing the student's understanding of Spearman's rank correlation coefficient (3.2) and testing the hypothesis that a correlation is zero (3.3).

In part (a) we are testing if the students can use the given formulae to obtain the product Spearman's rank correlation coefficient. Students who forgot to rank the data were able to gain only the 3rd M1.

Part (b) requires the students to carry out a hypothesis test. They need to give the hypotheses, write down the critical value, to the minimum accuracy used in the tables and draw a correct conclusion. The main error that occurred was not writing the hypotheses in terms of Rho.

Part (c) proved to be one of the more challenging parts of the paper. On the whole students knew that the correlation coefficient would increase but were unable to give a complete reason. The required reason consisted of two parts. The first was to realise that the $\sum d^2$ would decrease although we allowed d or d^2 decrease for D and E. The second part was to indicate that the ranks of D and E would be the same.
The most common explanation given that did not have the required detail was "the judges are in more agreement therefore Spearman's rank correlation coefficient would decrease".

Mark Scheme

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AOs</th></tr>
<tr><td>1(a)</td><td><table>
<tr><th>Skater
Judge</th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th><th>H</th></tr>
<tr><td>1</td><td>2</td><td>3</td><td>1</td><td>5</td><td>4</td><td>6</td><td>7</td><td>8</td></tr>
<tr><td>2</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>8</td><td>7</td></tr>
<tr><td>d^2</td><td>1</td><td>1</td><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
</table></td><td>M1
dM1</td><td>1.1b
1.1b</td></tr>
<tr><td></td><td>$\sum d^2 = 10$</td><td></td><td></td></tr>
<tr><td></td><td>$r_s = 1 - \frac{6\times"10"}{8(64-1)}$</td><td>M1</td><td>1.1b</td></tr>
<tr><td></td><td>$r_s = 0.8809...$ awrt 0.881</td><td>A1</td><td>1.1b</td></tr>
<tr><td></td><td></td><td>(4)</td><td></td></tr>
<tr><td>(b)</td><td>$\text{H}_0 : \rho = 0 \quad \text{H}_1 : \rho > 0$</td><td>B1</td><td>2.5</td></tr>
<tr><td></td><td>Critical Value $\rho = 0.8333$</td><td>B1</td><td>1.1b</td></tr>
<tr><td></td><td>$r_s = 0.8809$ lies in the critical region/reject H_0/significant</td><td>M1</td><td>2.1</td></tr>
<tr><td></td><td>There is evidence that the two judges are in agreement.</td><td>A1cso</td><td>2.2b</td></tr>
<tr><td></td><td></td><td>(4)</td><td></td></tr>
<tr><td>(c)</td><td>The $\sum d^2$ will decrease since the new rankings given by Judge 1 are now the same as the rankings given by Judge 2 for Skater D and E whereas previously they were different</td><td>M1</td><td>2.4</td></tr>
<tr><td></td><td>therefore Spearman's rank correlation coefficient will increase</td><td>A1</td><td>2.2a</td></tr>
<tr><td></td><td></td><td>(2)</td><td></td></tr>
<tr><td colspan="4" align="right">(10 marks)</td></tr>
<tr><th colspan="4">Notes</th></tr>
<tr><td colspan="4">(a) M1: For an attempt to rank at least one row (at least 4 correct)
dM1: dep on previous M mark being awarded. For an attempt at d or d^2 row for their ranks.
M1: for use of $1-\frac{6\times"\text{their} \sum d^2"}{8(64-1)}$ Allow if not ranked $\sum d^2 = 85$
A1: awrt 0.881
(b) B1: Both hypotheses stated in terms of ρ
B1: for correct critical value. Allow even if 2 tail test (sign must match their r_s)
M1: for comparing their 0.881 with "their 0.8333"
A1cso: All previous marks awarded. For a correct contextual conclusion with no contradictions seen

(c) M1: For a correct explanation to support their answer given.
$\sum d^2$ decreases or d/d^2 decreases for D and E
and idea of same rankings eg d^2 will reduce by 2 . Do not allow d^2 will reduce by 1
A1: for a correct deduction from the information. Allow closer to 1.</td></tr>
</table>

Student Response A

a) $1 - \frac{6\sum d^2}{n(n^2-1)}$ $n=8$

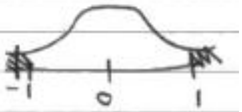
$\sum d^2 = (-2)^2 + (1)^2 + 5^2 + (-2)^2 + (1)^2 + 5^2 + 5^2 + \cancel{0^2}$

85

$\frac{6(85)}{8(63)} = \frac{510}{504} = 1.011904762$

$1 - \text{ANS} = -0.01190$

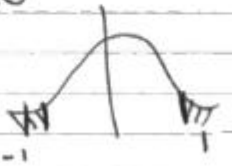
b) H_0: judges are in agreemnt
H_1: judges are not in agreement



siglevel 0.01 CV TS

$0.01 > -0.01190$

$\therefore$ accept H_1
enough evidence to ~~suggests~~ reject H_0



c)

71	70	72	63	62	61	57	53
73	71	67	64	62	56	52	53

The $\sum d^2$ value will be smaller meaning the $\frac{6\sum d^2}{n(n^2-1)}$ value will be smaller therefore meaning the Spearman rank will be lager

1/10

Examiner Comments

In part (a) there has been no attempt to rank, M0. The second Method mark is dependent on the first being awarded, M0. They have then used the correct formula with $n = 8$ and their sum of d^2 M1 but their answer is incorrect.
M0dM0M1A0

In part (b) the hypotheses have not been given in terms of Rho (they are also the wrong way round) No critical value is found. No critical value has been found so a comparison cannot be made.
B0B0M0A0cso

In part (c) they have realised that the sum of d^2 will decrease but not got the idea that this is because the rankings are the same.
M0A0

Student Response B

a)

Skater	RANK J1	J2	d	d^2
A	2	1	1	1
B	3	2	1	1
C	1	3	−2	4
D	5	4	0 1	1
E	4	5	0 −1	1
F	6	6	0	0
G	7	8	−1	1
H	8	7	1	1

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} \qquad n = 8$$

$$= 1 - \frac{6 \times 10}{8(8^2-1)}$$

$$r_s = 0.8809$$

b) H_0: the judges are in agreement
H_1: the judges are NOT in agreement

0.01 sig lvl $n = 8$

∴ critical value = 0.8333

$0.8809 > 0.8333$

∴ there is insufficient evidence to reject H_0 suggesting the judges are in agreement.

c) r_s would decrease as switching the scores flipps their ranks, lining them up with J2s ranks. thus d and ∴ d^2 would be lower, decreasing the value abstracted from 1 in $r_s = 1 - \frac{\Sigma d^2}{n(n^2-1)}$ and increasing r_s.

5/10

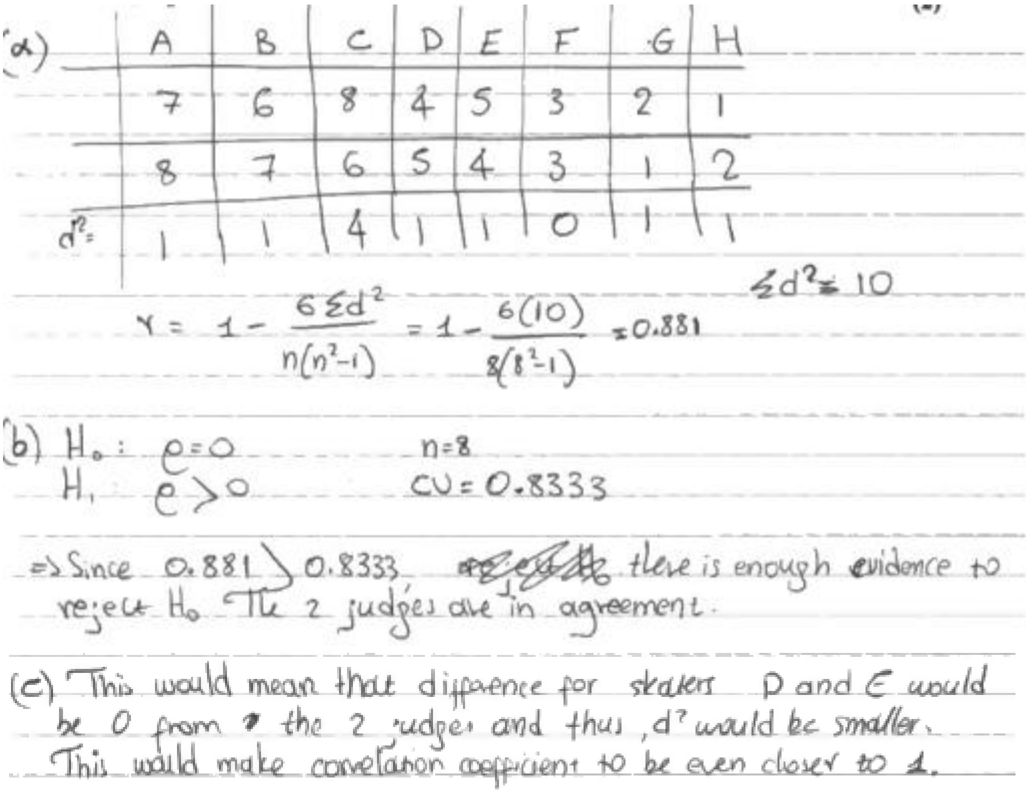
Examiner Comments

In part (a) the correct method has been used to find the awrt 0.881
M1dM1M1A1

In part (b) the hypotheses have not been given in terms of Rho (they are also the wrong way round). The Critical value is correct but there is an incorrect comparison as insufficient evidence to reject H_0 means H_0 is accepted. The final mark cannot be awarded as A0 follows M0.
B0B1M0A0cso

In part (c) there is the idea of the same rankings however, although they have said d^2 decreases it is the sum of d^2 that decreases although we allowed decreases for D and E
M0A0

Student Response C



(a)

	A	B	C	D	E	F	G	H
	7	6	8	4	5	3	2	1
	8	7	6	5	4	3	1	2
$d^2 =$	1	1	4	1	1	0	1	1

$\Sigma d^2 = 10$

$$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6(10)}{8(8^2-1)} = 0.881$$

(b) $H_0 : \rho = 0$ $n=8$

$H_1 \quad \rho > 0$ $CV = 0.8333$

$\Rightarrow$ Since $0.881 > 0.8333$ there is enough evidence to reject H_0. The 2 judges are in agreement.

(c) This would mean that difference for skaters D and E would be 0 from the 2 judges and thus, d^2 would be smaller. This would make correlation coefficient to be even closer to 1.

10/10

Examiner Comments

In part (a) the correct value is found to the appropriate degree of accuracy.
M1dM1M1A1

In part (b) the hypotheses have not been given in terms of Rho. The Critical value is correct and there is a correct comparison as H_0 is rejected. This is followed by a correct contextual statement.
B1B1M1A1cso

In part (c) stating the difference for *D* and *E* are 0 is equivalent to the rankings are the same and thus d^2 would be smaller is equivalent to d^2decreases for *D* and *E*. Closer to 1 is accepted as the Spearman's rank correlation coefficient will increase.
M1 A1

Exemplar Question 2

back to Contents Page

2. Lloyd regularly takes a break from work to go to the local cafe. The amount of time Lloyd waits to be served, in minutes, is modelled by the continuous random variable T, having probability density function

$$\mathrm{f}(t) = \begin{cases} \dfrac{t}{120} & 4 \leqslant t \leqslant 16 \\ 0 & \text{otherwise} \end{cases}$$

(*a*) Show that the cumulative distribution function is given by

$$\mathrm{F}(t) = \begin{cases} 0 & t < 4 \\ \dfrac{t^2}{240} - c & 4 \leqslant t \leqslant 16 \\ 1 & t > 16 \end{cases}$$

where the value of c is to be found.

(2)

(*b*) Find the exact probability that the amount of time Lloyd waits to be served is between 5 and 10 minutes.

(2)

(*c*) Find the median of T.

(2)

(*d*) Find the value of k such that

$$\mathrm{P}(T < k) = \frac{2}{3}\,\mathrm{P}(T > k)$$

giving your answer to 3 significant figures.

(3)

(Total for Question 2 is 9 marks)

Mean Score 6.3 out of 9

Examiner Comments

This question is assessing the student's understanding of probability density functions and cumulative distributions, their relationship and how to use them to find probabilities and the median (2.1, 2.2 and 2.3).

Part (a) was a "show that" question which means that just stating c is $\frac{1}{15}$ is not enough. A fully correct method must also be shown.

Parts (b) and (c) require the use of standard techniques.

Part (d) required the mathematical problem to be translated into an equation which could then be solved using any method.

Mark Scheme

Question	Scheme	Marks	AOs
2(a)	$\int \frac{t}{120}\mathrm{d}t = \frac{t^2}{240}$ and use of F(4) = 0 or F(16) = 1 or limits of t and 4	M1	2.1
	or attempt at area of trapezium allow 1 mistake. $\frac{1}{2}\times(t-4)\left(\frac{4}{120}+\frac{t}{120}\right)$		
	$=\frac{t^2}{240}-\frac{1}{15}$	A1	1.1b
		(2)	
(b)	F(10) – F(5) = $\frac{100}{240}-"c"-\frac{25}{240}+"c"$	M1	1.1b
	$=\frac{5}{16}$	A1	1.1b
		(2)	
(c)	$\frac{m^2}{240}-\frac{1}{15}=0.5$	M1	1.1b
	$m=11.66....$ awrt 11.7	A1	1.1b
		(2)	
(d)	$\mathrm{F}(k)=\frac{2}{3}\left(1-\mathrm{F}(k)\right)$ or $\int_4^k \frac{t}{120}\mathrm{d}t=\frac{2}{3}\int_k^{16}\frac{t}{120}\mathrm{d}t$	M1	3.1a
	$\frac{k^2}{240}-"\frac{1}{15}"=\frac{2}{3}\left(1-\left(\frac{k^2}{240}-"\frac{1}{15}"\right)\right)$ or $\frac{k^2}{240}-\frac{1}{15}=\frac{2}{3}\times\left(\frac{16}{15}-\frac{k^2}{240}\right)$	dM1	1.1b
	$\frac{k^2}{144}=\frac{7}{9}$		
	$k=\sqrt{112}$ or awrt 10.6	A1	1.1b
	Alternative		
	Let $\mathrm{P}(T<k)=p$ then $p=\frac{2}{3}(1-p)\ \therefore p=\frac{2}{5}$	(M1)	
	$\frac{k^2}{240}-\frac{1}{15}=\frac{2}{5}$	(dM1)	
	$k=\sqrt{112}$ or awrt 10.6	(A1)	
		(3)	
			(9 marks)

Notes

(a) **M1:** for attempting to integrate and a correct method

A1: $=\frac{t^2}{240}-\frac{1}{15}$ or $=\frac{t^2}{240}-0.0667$

(b) **M1:** writing or using F(10) – F(5)

A1: awrt $\frac{5}{16}$ or 0.3125 or exact equivalent

(c) **M1:** setting their F(t) = 0.5

A1: awrt 11.7 or $2\sqrt{34}$ or exact equivalent

(d) **M1:** Setting up a correct equation to solve the mathematical problem or setting up correct equation to find *p* and an attempt to solve

dM1: attempted to integrate and limits substituted or using “Their F(k)” = “their p”

A1: $\sqrt{112}$ or awrt 10.6

Student Response A

a) $\int_4^x \frac{t}{120} = \left[\frac{t^2}{120\times 2}\right]_4^x = \left[\frac{t^2}{240}\right]_4^x$

~~$\left[\frac{t^2}{240}\right]_4^{16} = \left(\frac{16^2}{240}\right) - \left(\frac{4^2}{240}\right) + c = 1$~~

~~$\frac{16}{15} - \frac{1}{15} + c = 1$~~

~~$1 + c = 1$~~

$c = 0$

b) $\int_5^{10} \frac{t}{120} = \left[\frac{t^2}{240}\right]_5^{10}$

$\left(\frac{10^2}{240}\right) - \left(\frac{5^2}{240}\right) = \frac{5}{12} - \frac{5}{48} = \frac{5}{16}$

c) $\frac{t^2}{240} = 0.5$

$t^2 = 120$

$t = 2\sqrt{30}$

d) $P(5/16 < k) = \frac{2}{3} P(\cancel{5}\, 5/16 > k)$

$3P(5/16 < k) = 2P(5/16 > k)$

4/9

Examiner Comments

In part (a) the integration is shown and the limits of x and 4 are used (we allow x instead of t) leading to the incorrect value of c being found.
M1A0

In part (b) the correct answer is given.
M1A1

In part (c) they have put their $\mathrm{F}(t) = 0.5$ since their value of c is zero.
M1A0

In part (d) a correct equation has been not set up
M0M0A0

Student Response B

(3)

(a) when $t \leq 4$ the probability of this must be zero. so area under graph when $t < 4$ is zero.

area under $\frac{t}{120}$ graph.

$$\int_4^x \left(\frac{t}{120}\right) dt = \left[\frac{t^2}{240}\right]_4^x = \left(\frac{x^2}{240}\right) - \frac{4^2}{240} = \frac{x^2}{240} - \frac{1}{15}$$

$C = \frac{1}{15}$

(c) $\frac{x^2}{240} - \frac{1}{15} = 0.5$

$\frac{x^2}{240} = \frac{17}{30}$ $\quad x^2 = 136$ $\quad x = 2\sqrt{34}$

median value $= 2\sqrt{34}$

(b) ~~$\int_5^{10} \frac{t}{120}$~~

$$\left(\frac{10^2}{240} - \frac{1}{15}\right) - \left(\frac{5^2}{240} - \frac{1}{15}\right) = 0.3125$$

$P(5 < t < 10) = 0.3125$

(d)

$k = t - 4$ $\quad k + 4 = t$

6/9

Examiner Comments

In part (a) the integration is shown and the limits of *t* and 4 are used leading to the correct value of *c* being found.
M1A1

Part (b) is after part (c) and the correct answer is given.
M1A1

In part (c) the correct answer is given.
M1A1

In part (d) no equation has been set up
M0M0A0

Student Response C

(a) $\int_4^t \frac{1}{120\cancel{0}} t \; dt = \left[\frac{1}{240} t^2\right]_4^t$

$= \frac{1}{240} t^2 - \frac{(4)^2}{240}$

$= \frac{t^2}{240} - \frac{1}{15}$

$$F(t) = \begin{cases} 0 & t < 4 \\ \frac{t^2}{240} - \frac{1}{15} & 4 \leq t \leq 16 \\ 1 & t > 16 \end{cases}$$

$c = \frac{1}{15}$

(b) $P(5 \leq t \leq 10) = P(t \leq 10) - P(t \leq 5)$

$= \left(\frac{10^2}{240} - \frac{1}{15}\right) - \left(\frac{5^2}{240} - \frac{1}{15}\right)$

$= \frac{7}{20} - \frac{3}{80}$

$= \frac{5}{16}$

$= \underline{0.3125}$

(c) $F(m) = \frac{m^2}{240} - \frac{1}{15} = \cancel{0.25}\ \frac{1}{2}$

$\frac{2m^2}{240} - \frac{2}{15} = 1$

$\frac{m^2}{120} - \frac{2}{15} = 1$

$m^2 = \left(1 + \frac{2}{15}\right) 120$

$m^2 = 136$

$m = \sqrt{136} =$ 11.7 minutes (3sf)

(d) $P(T < k) = \frac{2}{3} P(T > k)$

$$P(T<k) = \frac{2}{3}\left(1 - P(T \leq k)\right)$$

$$= \frac{2}{3} - \frac{2}{3} P(T \leq k)$$

$$P(T<k) = \frac{k^2}{240} - \frac{1}{15}$$

$$\frac{k^2}{240} - \frac{1}{15} = \frac{2}{3} - \frac{2}{3}\left(\frac{k^2}{240} - \frac{1}{15}\right)$$

$$= \frac{2}{3} - \frac{2k^2}{720} + \frac{2}{45}$$

$$\frac{3k^2}{720} = \frac{32}{45} - \frac{2k^2}{720}$$

$$\frac{5k^2}{720} = \frac{32}{45}$$

$$5k^2 = 512$$

$$k^2 = 102.4$$

$$k = \sqrt{102.4} = 10.1192885|$$

~~k = 10.12 minutes~~

k = 10.1 minutes

8/9

Examiner Comments

In part (a) the integration is shown and the limits of t and 4 are used leading to the correct value of c being found.
M1A1

In part (b) the correct answer is found.
M1A1

In part (c) the correct answer is found.
M1A1

In part (d) a correct equation has been given. A mistake has been made when rearranging the equation to isolate k^2.
M1M1A0

Exemplar Question 3

back to Contents Page

3. Two students, Jim and Dora, collected data on the mean annual rainfall, w cm, and the annual yield of leeks, l tonnes per hectare, for 10 years.

Jim summarised the data as follows

$$S_{wl} = 42.786 \qquad S_{ww} = 9936.9 \qquad \sum l^2 = 26.2326 \qquad \sum l = 16.06$$

(*a*) Find the product moment correlation coefficient between l and w

(2)

Dora decided to code the data first using $s = w - 6$ and $t = l - 20$

(*b*) Write down the value of the product moment correlation coefficient between s and t. Give a justification for your answer.

(1)

Dora calculates the equation of the regression line of t on s to be $t = 0.00431s - 18.87$

(*c*) Find the equation of the regression line of l on w in the form $l = a + bw$, giving the values of a and b to 3 significant figures.

(3)

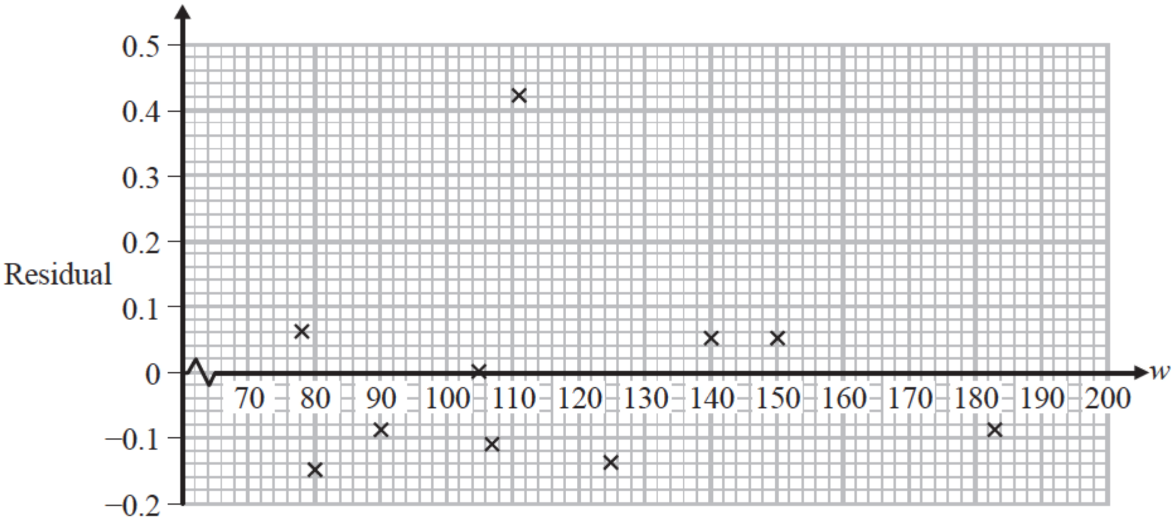
(*d*) Use your equation to estimate the yield of leeks when w is 100 cm.

(1)

(*e*) Calculate the residual sum of squares.

(2)

The graph shows the residual for each value of l



(*f*) (i) State whether this graph suggests that the use of a linear regression model is suitable for these data. Give a reason for your answer.

(ii) Other than collecting more data, suggest how to improve the fit of the model in part (*c*) to the data.

(2)

(Total for Question 3 is 11 marks)

Mean Score 8.6 out of 11

Examiner Comments

This question is assessing the student's understanding of least squares regression (1.1), residuals (1.2) and the product moment correlation coefficient (3.2)

Part (a) required the students to use the formulae to find the product moment correlation coefficient.

In part (b) whilst most students realised that the product moment correlation coefficient would be the same as part (a) a minority of students either forgot or were unable to give a suitable reason.

Part (c) required the substitution of $s = w - 6$ and $t = l - 20$ into the equation $t = 0.00431s - 18.87$ This was the most popular and successful. The most commonly used alternative method was to go back to the original summary data to find the value of b and then use a particular value for s to try and find the value of a. This was rather long winded and generally unsuccessful.

In part (e) either $w = 100$ or $s = 94$ needs to be substituted into the relevant equation.

Part (f) was often left blank. Students should remember that leaving a part unanswered will definitely gain zero marks but if an attempt is made there is always a chance of scoring some marks. A correct answer to both parts was rarely seen. The most common error in part (i) was to state that "all the points are close to zero" which may be true but does not necessarily mean a linear regression model is suitable. They need to be randomly scattered. Part (ii) required the recognition that there was an outlier which could be removed in order to improve the fit of the model.

Mark Scheme

Question	Scheme	Marks	AOs
3(a)	$\left[S_{ll}=26.2326-\frac{16.06^2}{10}=0.44024\right]$		
	$r=\frac{42.786}{\sqrt{9936.9\times"0.44024"}}$	M1	1.1b
	$r=0.64689...$ awrt 0.647	A1	1.1b
		(2)	
(b)	$"0.647"$ coding has no effect on the pmcc	B1ft	1.1b
		(1)	
(c)	$l-20=0.00431(w-6)-18.87$	M1	3.1a
	$l=0.00431w+....$	M1	1.1b
	$l=0.00431w+1.10414$	A1	1.1b
		(3)	
(d)	$l=0.00431\times100+1.10=1.53$	B1ft	3.4
		(1)	
(e)	$\text{RSS}="0.44024"-\frac{(42.786)^2}{9936.9}$ or $"0.44024"\left(1-"0.647"^2\right)$	M1	1.1b
	RSS = 0.2560	A1	1.1b
		(2)	
(f)	(i) The points appear **randomly** scattered above and below zero giving us no reason to doubt the suitability of the linear model.	B1	3.5a
	(ii) There is a possible outlier that could be removed (and the regression line recalculated).	B1	3.5c
		(2)	
(11 marks)			

Notes
(a) M1: For a complete correct method to find *r*
A1: for awrt 0.647
(b) B1ft: stating their answer to part (a) and a correct reason
(c) M1: for use of a correct model. i.e. a correct expression for *b*
M1: for use of a correct model i.e. a correct expression (ft) for *a*
A1: for correct model $l=0.00431w+1.10$ with awrt 0.00431 and awrt 1.10
(d) B1ft: correct answer using their equation and *w* = 100 or using $t=0.00431s-18.87$ and $s=94$ Allow awrt 1.53/1.54
(e) M1: for a correct expression for RSS
A1: awrt 0.256
(f) B1: For explaining why the model may be suitable. Allow randomly scattered around *w* (*x*) axis. Do not allow most residuals close to zero or not suitable as not randomly scattered.
B1: For explaining how the fit of the model might be improved.

Student Response A

(a) $S_{ll} = \sum l^2 - \frac{(\sum l)^2}{n} = 26.2326 - \frac{(16.06)^2}{10}$

$= 26.2326 - 25.79236 = 0.44024$

$r = \frac{S_{wl}}{\sqrt{S_{ll} \times S_{ww}}} = \frac{42.786}{\sqrt{9936.9 \times 0.44024}} = 0.6468915486$

$= 0.6469(4sf)$

(b) —

(c) $l = 0.004315 - 18.57$

$l - 20 = 0.00431(w - 6) - 18.57$

$l = 0.00431w - 0.02586 - 18.57 + 20$

$l = 0.00431w - 1.15586.$

4/11

Examiner Comments
In part (a) an awrt 0.647 is given.
M1A1

In part (b) no answer is given.
B0ft

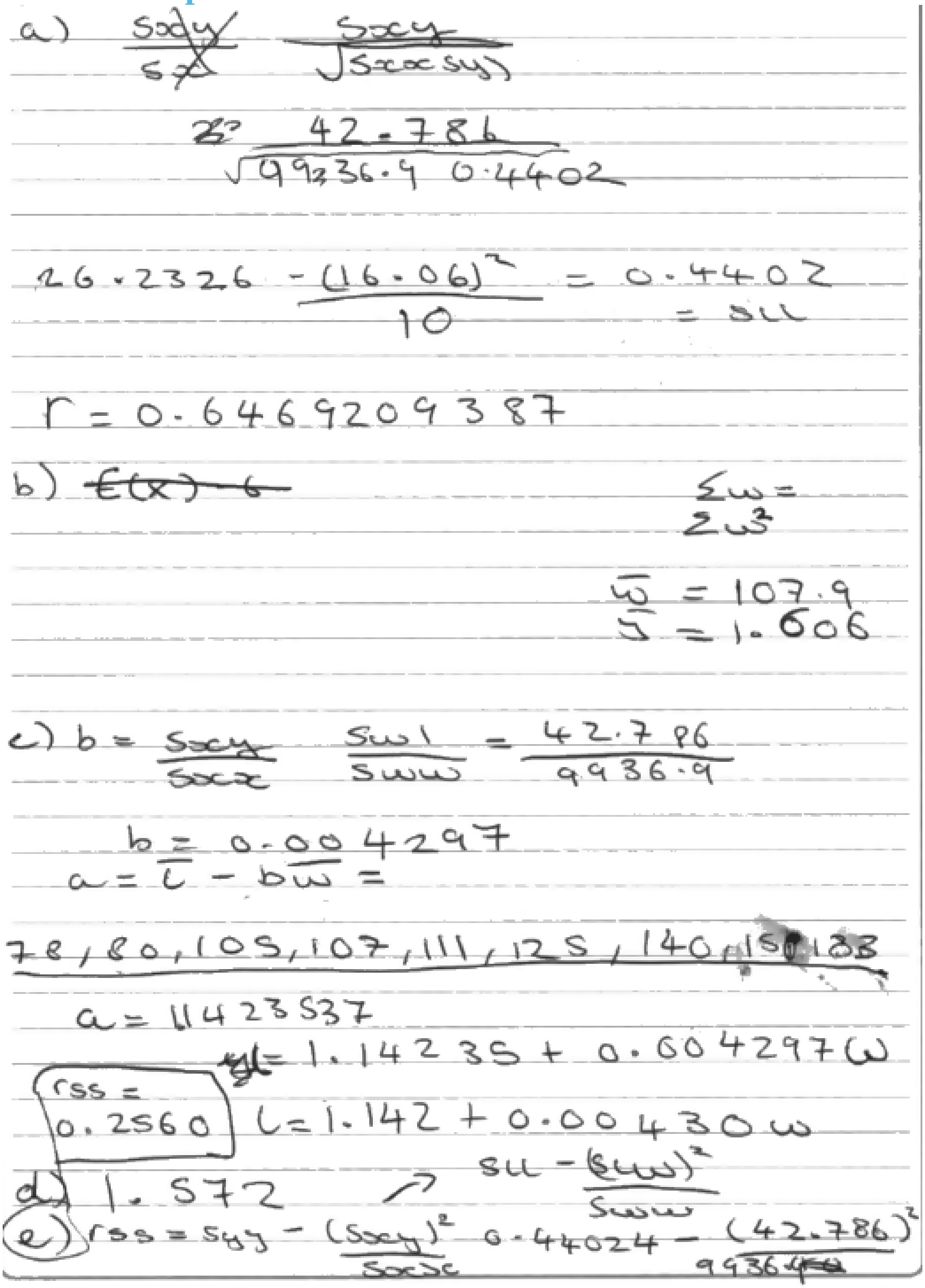
In part (c) the correct model has been used resulting in the equation of the regression line being in the form *l* = 0.00431 +…. We allow *l* = 0.00431 –…. The final equation of the regression line is incorrect
M1M1A0

In part (d) no answer is given.
B0ft

In part (e) no answer is given.
M0A0

In part (f) no answer is given.
B0B0

Student Response B



a) $\cancel{\frac{Sxdy}{Sx}}$ $\frac{Sxy}{\sqrt{Sxx\,Syy}}$

$\cancel{z}$ $\frac{42.786}{\sqrt{9936.4 \quad 0.4402}}$

$26.2326 - \frac{(16.06)^2}{10} = 0.4402$

$= S_{ll}$

$r = 0.6469209387$

b) ~~$E(x) - 6$~~

$\Sigma w =$

Σw^2

$\bar{w} = 107.9$

$\bar{l} = 1.606$

c) $b = \frac{Sxy}{Sxx}$ $\frac{Swl}{Sww} = \frac{42.786}{9936.9}$

$b = 0.004297$

$a = \bar{l} - b\bar{w} =$

78, 80, 105, 107, 111, 125, 140, 15[illegible] 188

$a = 11423537$

$\cancel{y} = 1.14235 + 0.004297w$

$\boxed{rss = 0.2560}$ $l = 1.142 + 0.00430w$

$S_{ll} - \frac{(S_{lw})^2}{S_{ww}}$

d) 1.572 ↗

e) $rss = Syy - \frac{(Sxy)^2}{Sxx}$ $0.44024 - \frac{(42.786)^2}{9936.~~4~~}$

fi) this suggests the graph is not Suitable for a linear regression model as it doesn't have an elliptical Shape

fii) remove value (111, 0.42) from the data as it ~~will~~ is an outlier and will reduce Pearsons product moment correlation

7/11

Examiner Comments

In part (a) an awrt 0.647 is given.
M1A1

In part (b) no value for the pmcc is given.
B0ft

In part (c) there is a correct method shown for finding the value of *b*. However an incorrect method has been used to find the value of *a* and the answer is incorrect.
M1M0A0

In part (d) 1.572 is correct using their value of *a* and their value of *b*.
B1ft

In part (e) a correct formula has been used leading to the correct RSS value.
M1A1

In part (f) (i) there is the suggestion that the model is not suitable and there is no reference to the points being randomly scattered.
B0

In (ii) there is the suggestion that the outlier be removed. This is all that is needed for **B1**.

Student Response C

ⓐ. $S_{ll} = \Sigma l^2 - \frac{(\Sigma l)^2}{n}$

$= 26.2326 - \frac{16.06^2}{10}$

$\doteq 0.44024$

$r = \frac{S_{lw}}{\sqrt{S_{ll} \times S_{ww}}} = \frac{42.786}{\sqrt{9936.9 \times 0.44024}} = 0.6468913486$

$= 0.647$

ⓑ. $r = 0.647$

A linear coding does not change the product moment correlation coefficient.

ⓒ. $s = w - 6 \quad t = l - 20$

$t = 0.00431s - 18.87$

$l - 20 = 0.00431(w-6) - 18.87$

$l = 0.00431w + 20 - 0.00431 \times 6 - 18.87$

$l = 0.00431w + 20 - 0.02586 - 18.87$

$l = 0.00431w + 1.10414$

$\therefore l = 0.00431w + 1.10$

$\therefore a = 1.10, \quad b = 0.00431$

ⓓ. $l = 0.00431w + 1.10$

$w = 100$

$l = 0.00431 \times 100 + 1.10$

$l = 0.431 + 1.10$

$l = 1.531$

$l = 1.53$ tonnes per hectare

ⓔ. $RSS = S_{ll}(1 - r^2)$

$= 0.44024 \times (1 - 0.6468913486)$

$= 0.1554524646$

$= 0.155$

P.T.O

ⓕ. (I). A linear model may be suitable as the residuals are randomly distributed about 0

(II). Calculate the summary statistics for s and t and use those to work out the equation of the regression line

9/11

Examiner Comments

In part (a) an awrt 0.647 is given.
M1A1

In part (b) the correct pmcc is given with a correct reason.
B1ft

In part (c) the correct model is given.
M1M1A1

In part (d) the correct answer is given.
B1ft

In part (e) a correct formula has been written down and an attempt to use it has been made. As the formula has been written down we can treat the missing squared as a slip. If the formula had not been written down this would have gained M0.
M1A0

In part (f) (i) they have the residuals are randomly scattered about 0.
B1

In (ii) there is no acceptable explanation on how to improve the model.
B0

Exemplar Question 4

back to Contents Page

The random variable X has a continuous uniform distribution over the interval $[5, a]$, where a is a constant.

Given that $\text{Var}(X) = \dfrac{27}{4}$

(*a*) show that $a = 14$

(3)

The continuous random variable Y has probability density function

$$\text{f}(y) = \begin{cases} \dfrac{1}{20}(2y - 3) & 2 \leqslant y \leqslant 6 \\ 0 & \text{otherwise} \end{cases}$$

The random variable $T = 3(X^2 + X) + 2Y$

(*b*) Show that $\text{E}(T) = \dfrac{9857}{30}$

(7)

(Total for Question 4 is 10 marks)

Mean Score 6.7 out of 10

Examiner Comments

This question is assessing the student's understanding of the continuous uniform distribution (2.3) and probability density functions (2.1) and how to find the mean (2.4).

Part (a) required the students to use the formula for the variance of a uniform distribution to set up and equation and then solve it. The answer is given so all working must be shown and in particular a reason must be given as to why 14 and not -4 was the required value.

Part (b) required some problem solving in order to work out the steps required to reach the given answer. They needed to find $\text{E}(X)$, $\text{E}(X^2)$, $\text{E}(Y)$ and use $\text{E}(aX) = a\text{E}(X)$.

Mark Scheme

Question	Scheme	Marks	AOs
4(a)	$\frac{1}{12}(a-5)^2 = \frac{27}{4}$	M1	3.1a
	$(a-5)^2 = 81$		
	$a-5=9$ **or** $a-5=-9$	A1	1.1b
	$\therefore$ since $a>5$ $a = 14$*	A1cso*	2.2a
		(3)	
(b)	Correct method for E(Y), E(X) and E(X^2) **or** E(Y) and E(X^2+X)	M1	3.1a
	$\mathrm{E}(Y) = \int_2^6 \frac{1}{20} y(2y-3)\,\mathrm{d}y$	M1	1.1b
	$= \frac{68}{15}$	A1	1.1b
	$\mathrm{E}(X) = \frac{5+14}{2}$ or 9.5 **and** $\frac{27}{4} = \mathrm{E}(X^2) - 9.5^2$ or $\int_5^{14} \frac{x^2}{9}\mathrm{d}x$ **or** $\int_5^{14}\left(\frac{x^3}{9}+\frac{x^2}{9}\right)\mathrm{d}x$ or $3\int_5^{14}\left(\frac{x^3}{9}+\frac{x^2}{9}\right)\mathrm{d}x$	M1	1.1b
	$\mathrm{E}(X^2) = 97$ and $\mathrm{E}(X) = 9.5$ **or** $\mathrm{E}(X^2+X) = 106.5$ **or** $3\mathrm{E}(X^2+X) = 319.5$	A1	1.1b
	$\mathrm{E}(T) = 3\times\text{"}97\text{"} + 3\times\text{"}9.5\text{"} + 2\times\frac{68}{15}$ oe	M1	1.1b
	$\mathrm{E}(T) = \frac{9857}{30}$ *	A1*cso	2.1
		(7)	
			(10 marks)

Notes

(a) **M1:** translating a problem in mathematical contexts into a correct equation. Allow

$$\frac{a^3-125}{3(a-5)} - \left(\frac{a+5}{2}\right)^2 = \frac{27}{4}$$

A1: for $a-5=9$ or $a-5=-9$ or $a^2-10a-56=0$ or $a^3-15a^2-6a+280=0$

A1cso*: concluding it is 14 giving a reason why -4 is rejected

(b) **M1**: For a complete method to solve the problem

M1: For an attempt at E(Y)

A1: $= \frac{68}{15}$ or awrt 4.53

M1: For an attempt at E(X) and E(X^2) or $\mathrm{E}(X^2+X)$ or $3\,\mathrm{E}(X^2+X)$

Some sort of working must be seen for E(X^2) eg $\frac{27}{4}$ = E(X^2) – E(X)^2.

Allow Var(X) = E(X^2) – E(X)^2 leading to = E(X^2)

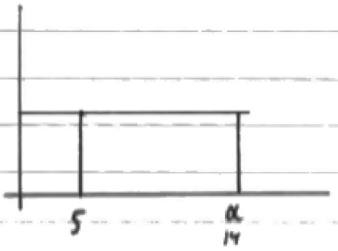
A1: 319.5

M1: Method for finding E(T) ft their values

A1*cso: Fully correct solution no errors, must have $\mathrm{E}(T) = \frac{9857}{30}$ *

Student Response A

(a)



$\int x f x$

$\frac{1}{12}(b-a)^2 = \frac{1}{12}(a-5)^2 = \frac{1}{12}a^2 - 10a + 25 = \frac{27}{4}$

$a^2 - 10a - 56 = 0$

$(a - 14)(a + 4) = 0$

$a = 14$ or $a = -4$

α cant be negative so $\alpha = 14$

(b) $T = 3x^2 + 3X + 2y$

$(14 - 5) \times p = 1$

$9p = 1$

$p = \frac{1}{9}$ ~~$E(X) = x\,P(X=x)$~~

$E(X) = \frac{1}{2}(a+b)$. $\frac{1}{2}(5 + 14) = \frac{19}{2}$

$E(X^2) = \sum_1 x^2$

$\int_5^{14} x\frac{1}{9} = \left(\frac{1}{18}x^2\right)_5^{14} = \frac{19}{2}$

$3 \times \frac{19}{2} + 3 \times \frac{1}{9} +$

3/10

Examiner Comments

In part (a) a correct equation has been formed and then solved correctly. A reason has been given as to why –4 has been rejected.
M1M1A1

In part (b) there is no method for finding E(Y). E(X) is correct but an incorrect method has been used to find E(X^2). There is only a partial calculation for Var(T).
M0M0A0M0A0M0A0

Student Response B

(a). $\mathrm{Var}(X) = \frac{27}{4}$

$\frac{(a-5)^2}{12} = \frac{27}{4}$

$(a-5)^2 = \frac{27}{4} \times 12$

$(a-5)^2 = 27 \times 3$

$(a-5)^2 = 81$

$a - 5 = \pm 9$

$\therefore a = 9 + 5 = 14$

~~E(x~~ $\mathrm{Var}(X) = E(X^2) - \mu^2$

$\frac{27}{4} = E(X^2) - 14^2$

$E(X^2) = \frac{27}{4} + 14^2$

$= 202.75$

(B) $E(T) = E[3(X^2+X) + 2Y]$

$= E(3X^2 + 3X + 2Y)$

$= 3E(X^2) + 3E(X) + 2E(Y)$

$= 3\times 202.75 + 3\times 9.5 + 2\times\frac{272}{60}$

$= \frac{9857}{30}$

$E(X) = \frac{14+5}{2} = \frac{19}{2} = 9.5$

$E(y) = \int_2^6 \frac{4}{20}(2y-3)\,dy$

$= \frac{1}{20}\int (2y^2 - 3y)\,dy \Big|_2^6$

$= \frac{1}{20}\left[2\frac{1}{3}y^3 - 3\frac{1}{2}y^2\right]\Big|_2^6$

$= \frac{1}{20}\left(\frac{2}{3}y^3 - \frac{3}{2}y^2\right)\Big|_2^6$

$= \frac{1}{20}\left[(144-54) - \left(\frac{16}{3} - 6\right)\right]$

$= \frac{1}{20}(90 + 0.67)$

$= 4.5335$

$\left(90 - \frac{2}{3}\right)\times\frac{1}{20}$

$= \frac{\cancel{270\times2}\ 272}{3\times 20} = \frac{272}{60}$

$= \frac{540}{1160}$

5/10

Examiner Comments

In part (a) a correct equation has been formed and then solved correctly. Although –4 has been eliminated, no reason has been given as to why.

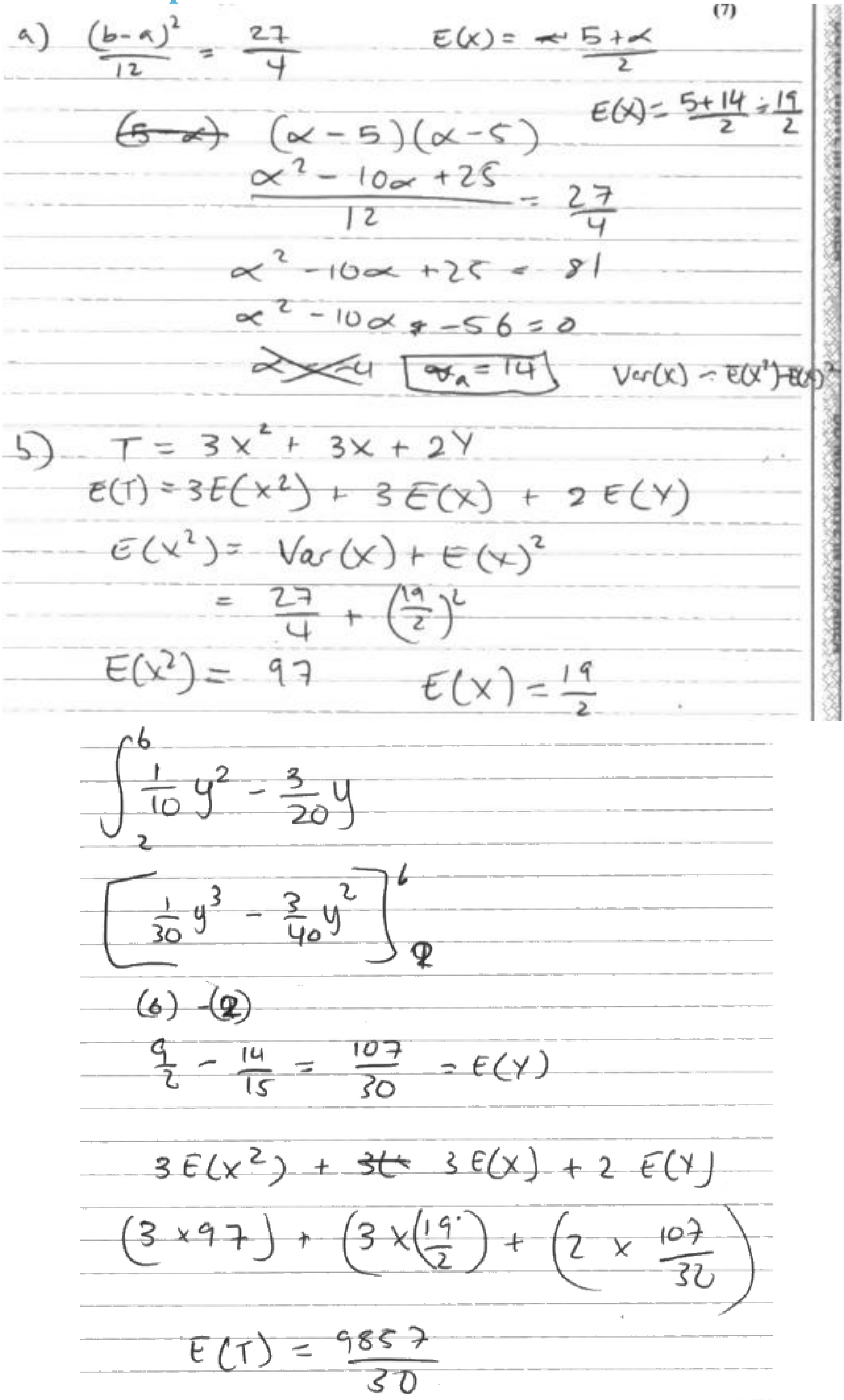
M1M1A0

In part (b) E(X) is correct. E(Y) is an awrt 4.53 An incorrect method has been used to find E(X^2) since they have not subtracted $[\mathrm{E}(X)]^2$.

A correct method for finding Var(T) has been used but the final A1cso requires all previous marks to have been awarded.

M0M1A1M0A0M1A0

Student Response C



(7)

a) $\frac{(b-a)^2}{12} = \frac{27}{4}$ $E(x) = \cancel{\alpha}\,\frac{5+\alpha}{2}$

~~(5 - α)~~ $(\alpha-5)(\alpha-5)$ $E(X) = \frac{5+14}{2} = \frac{19}{2}$

$\frac{\alpha^2 - 10\alpha + 25}{12} = \frac{27}{4}$

$\alpha^2 - 10\alpha + 25 = 81$

$\alpha^2 - 10\alpha \cancel{+} -56 = 0$

~~$\alpha = -4$~~ $\boxed{\cancel{\alpha}_a = 14}$ $Var(x) = E(x^2) - E(x)^2$

b) $T = 3X^2 + 3X + 2Y$

$E(T) = 3E(X^2) + 3E(X) + 2E(Y)$

$E(X^2) = Var(X) + E(X)^2$

$= \frac{27}{4} + \left(\frac{19}{2}\right)^2$

$E(X^2) = 97$ $E(X) = \frac{19}{2}$

$\int_2^6 \frac{1}{10}y^2 - \frac{3}{20}y$

$\left[\frac{1}{30}y^3 - \frac{3}{40}y^2\right]_2^6$

$(6) - (2)$

$\frac{9}{2} - \frac{14}{15} = \frac{107}{30} = E(Y)$

$3E(X^2) + \cancel{3E}\ 3E(X) + 2E(Y)$

$(3 \times 97) + \left(3 \times \left(\frac{19}{2}\right)\right) + \left(2 \times \frac{107}{30}\right)$

$E(T) = \frac{9857}{30}$

7/10

Examiner Comments

In part (a) a correct equation has been formed and then solved correctly. Although –4 has been eliminated, no reason has been given as to why.
M1M1A0

In part (b) E(X) and E(X^2) have been found correctly and a correct method for E(Y) has been used but the final answer is incorrect. It should be $\frac{1}{30}$ not $\frac{14}{15}$ when 2 was substituted.
A correct method for finding Var(T) has been used but the final A1cso requires all previous marks to have been awarded. The answer is given and had the student calculated their second to last line they would have seen it does not give the required answer.
M1M1A0M1A1M1A0

A Level Further Mathematics – Further Statistics 1 (9FM0 3B)

Exemplar Question 1

back to Contents Page

1. A chocolate manufacturer places special tokens in 2% of the bars it produces so that each bar contains at most one token. Anyone who collects 3 of these tokens can claim a prize.

Andreia buys a box of 40 bars of the chocolate.

(*a*) Find the probability that Andreia can claim a prize.

(2)

Barney intends to buy bars of the chocolate, one at a time, until he can claim a prize.

(*b*) Find the probability that Barney can claim a prize when he buys his 40th bar of chocolate.

(3)

(*c*) Find the expected number of bars that Barney must buy to claim a prize.

(1)

(Total for Question 1 is 6 marks)

Mean Score 4.65 out of 6

Examiner Comments

Part (a) was a straightforward opener and most students answered this correctly.

In part (b) many recognised that a negative binomial model was required, and this was often correct although some lost an accuracy mark for failing to give their answer correct to 3 significant figures.

The other common error here, and in part (c), was to try and use a geometric model. In the final part most could use the formula given in the formula booklet correctly, but some weren't totally sure about the notation and therefore which values to use.

Mark Scheme

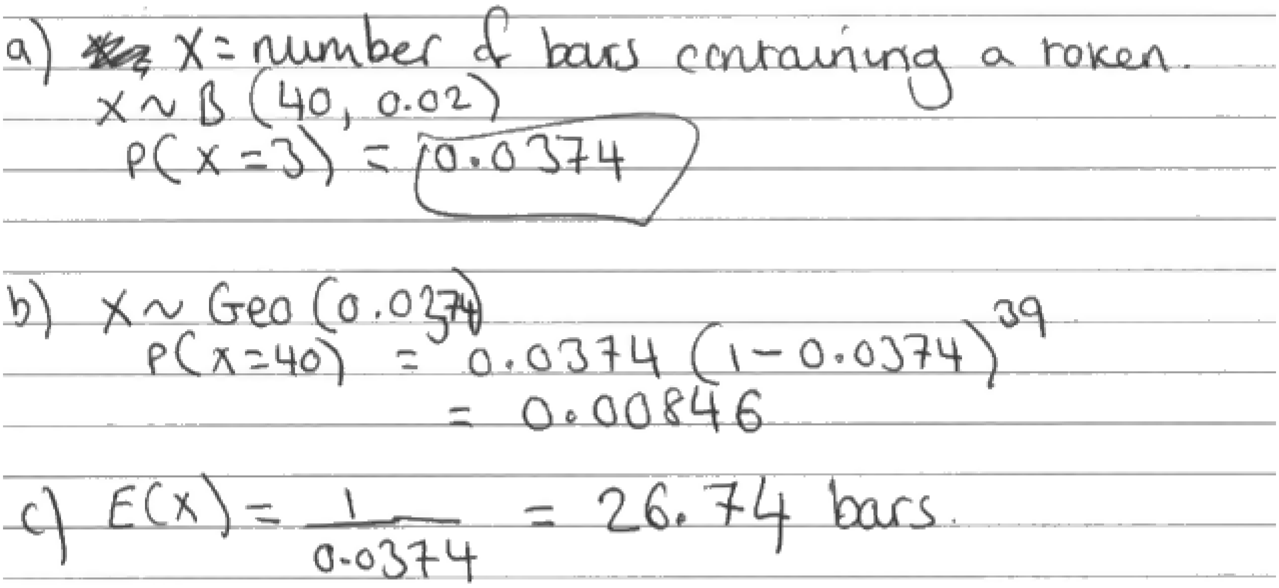
Qu	Scheme	Marks	AO
1(a)	[Let X = no. of prizes Andreia wins] $X \sim \mathrm{B}(40, 0.02)$	M1	3.3
	[Require $\mathrm{P}(X \geqslant 3) = 1 - \mathrm{P}(X \leqslant 2)$] $= 0.04567\ldots$ awrt **0.0457**	A1	1.1b
		(2)	
(b)	[Let Y = no. of the bar when Barney wins] $Y \sim \mathrm{NegBin}(3, 0.02)$	M1	3.3
	[$\mathrm{P}(Y = 40) =$] $\binom{39}{2} \times 0.02^2 \times 0.98^{37} \times 0.02$	M1	3.4
	$= 0.0028071\ldots$ awrt **0.00281**	A1	1.1b
		(3)	
(c)	$\mathrm{E}(Y) = \dfrac{3}{0.02} =$ **150**	B1	1.1b
		(1)	
		(6 marks)	

	Notes
(a)	M1 for selecting a suitable model i.e. $\mathrm{B}(40, p)$ where p is any probability Written or used, may be implied by a correct ans or 0.037429… from $\mathrm{P}(X = 3)$ A1 for awrt 0.0457 (correct answer only 2/2)
(b)	1st M1 for selecting a suitable model (NB(3, 0.02)) May be implied by a correct expression 2nd M1 for use of model to form a correct expression
SC	$\boldsymbol{p} \neq \mathbf{0.02}$ Allow prob of the form $\binom{39}{2} p^3(1-p)^{37}$ where $0 < p < 1$ scores M0M1 A1 for awrt 0.00281 (accept awrt 2.81×10^{-3}) [correct answer with no working scores 3/3]
(c)	B1 for 150

Examiner Comments

One of the assessment objectives being targeted in parts (a) and (b) requires the students to select a suitable model. A simple clear statement of the correct model [B(40, 0.02) in (a) and NegB (3, 0.02) in (b)] would score these method marks though, of course were also awarded for clear use of the correct model.

Student Response A



a) X = number of bars containing a token.

$X \sim B(40, 0.02)$

$P(X=3) = \boxed{0.0374}$

b) $X \sim Geo(0.0374)$

$P(X=40) = 0.0374(1-0.0374)^{39}$

$= 0.00846$

c) $E(X) = \frac{1}{0.0374} = 26.74$ bars.

1/6

Examiner Comments

In part (a) the correct binomial model is given (M1) but a number of candidates mis-interpreted the context like this and simply found P(X = 3) rather than P(X ⩾ 3) (A0).

In part (b) the correct model was not stated nor was there evidence of it’s use (M0A0). A number of students used their answer from part (a) as the probability here and a special case on the mark scheme enabled them to score 1 mark if a negative binomial model were being used but that is not the case here.

In part (c) this was not correct (B0).

Student Response B

(a) $X \sim B(40, 0.02)$

$P(X \leq 3) = 0.9918$

(b) $X \sim NB(\overset{3}{\cancel{40}}, 0.02)$

$P(X = 40) = {}^{39}C_2 (0.02)^3 (1-0.02)^{\cancel{38}\,37}$

$= 0.0028$

(c)

3/6

Examiner Comments

In part (a) the correct model is given (M1) but the answer is incorrect (A0).

In part (b) the correct model is stated (M1) and a correct expression is given (M1) but the answer is only given to 2s.f. not 3 and so the last mark was lost (A0).

In part (c) there is no attempt (B0).

Student Response C

the discrete random variable (DRV)

(a) Let X be the number of tokens found where
$X \sim B(3, 0.02)$

~~P(Andrea can claim prize) = P(Z=2) X 0.02~~
~~Where Z is a discrete random variable ~ B(39, 0.02)~~

$P(X \geq 3) = 1 - P(X \leq 2)$

$= 1 - 0.9543$

$= 0.0457$

(b) Let the descrete random variable Z be the 3 number of tokens barney finds, where $Z \sim NB([illegible], 0.02)$

~~P(Z = 40)~~

Let Y be the descrete random variable of the number of tokens barney finds in the first 39 bars, where
$Y \sim B(39, 0.02)$

$P(Z = 40) = P(Y = 2) \times 0.02$

(3sf)

$= 0.1404 \times 0.02 = 0.0028$

(c) ~~E(X) =~~
~~= 40 × 0.02~~

$\text{mean} = \frac{r}{p} = \frac{40}{0.02} = 2000$

5/6

Examiner Comments

In part (a) the model written at the start of the question is not correct but the following work is crossed out and the final answer presented is correct. We will give the mark for the correct model if a correct answer is given. (M1A1).

In part (b) the correct model is stated (M1) and used to give the correct answer to 3s.f. (M1A1).

In part (c) a number of students, such as this one, knew where to find the correct formula but were clearly not very confident about using it. An incorrect value of *r* is used and the answer is incorrect (B0).

Exemplar Question 2

back to Contents Page

2. Indre works on reception in an office and deals with all the telephone calls that arrive. Calls arrive randomly and, in a 4-hour morning shift, there are on average 80 calls.

(*a*) Using a suitable model, find the probability of more than 4 calls arriving in a particular 20-minute period one morning.

(3)

Indre is allowed 20 minutes of break time during each 4-hour morning shift, which she can take in 5-minute periods. When she takes a break, a machine records details of any call in the office that Indre has missed.

One morning Indre took her break time in 4 periods of 5 minutes each.

(*b*) Find the probability that in exactly 3 of these periods there were no calls.

(2)

On another occasion Indre took 1 break of 5 minutes and 1 break of 15 minutes.

(*c*) Find the probability that Indre missed exactly 1 call in each of these 2 breaks.

(3)

(Total for Question 2 is 8 marks)

Mean Score 6.50 out of 8

Examiner Comments

This was another well answered question with over 50% of the students scoring full marks. Most identified the Poisson distribution as a suitable model here and usually they had the correct mean too. A number of lost marks here over the interpretation of "more than 4 calls" and statements such as
$P(C > 4) = 1 - P(C \leqslant 3)$ were fairly common.

Part (b) was a little more demanding and whilst a number were able to use a Poisson distribution to obtain the probability of 0.189 they often did not realise that this was the parameter of a binomial distribution and answers of $0.189^3 \times (1-0.189)$ without the appropriate binomial coefficient were often seen. A few students rounded to 3 decimal places rather than 3 significant figures and, without sight of their full answer from their calculator, the examiners could not award the final mark. Whilst most identified two correct Poisson distributions e.g. $X \sim \text{Po}(\frac{5}{3})$ and $Y \sim \text{Po}(5)$ there were 3 common errors: some added the required probabilities instead of multiplying them together; others calculated $P(X = 1) \times P(Y = 0) + P(X = 0) \times P(Y = 1)$ and some used a $\text{Po}(\frac{5}{3} + 5)$ distribution.

Mark Scheme

Qu	Scheme	Marks	AO
2(a)	{Let C = no of calls in a 20 min period} $C \sim \text{Po}(\ldots)$	M1	3.3
	80 calls per 4-hour period gives $\frac{20}{3}$ per 20 mins i.e. $C \sim \text{Po}(\frac{20}{3})$	M1	3.4
	$[\text{P}(C > 4)] = 1 - \text{P}(C \leqslant 4)$		
	$= 0.79437\ldots$ awrt **0.794**	A1	1.1b
		(3)	
(b)	$\{X$ = no. of 5 min periods with no calls $\}$ $X \sim \text{B}(4, \text{e}^{-\frac{5}{3}})$	M1	3.3
	$\text{P}(X = 3) = 0.02186125\ldots$ awrt **0.0219**	A1	1.1b
		(2)	
(c)	P(exactly one call) $\text{e}^{-\frac{5}{3}} \times \frac{5}{3}$ or $\text{e}^{-5} \times 5$	M1	2.1
	P(exactly one call in each break) = $\left(\text{e}^{-\frac{5}{3}} \times \frac{5}{3}\right) \times \left(\text{e}^{-5} \times 5\right)$	M1	1.1b
	$= 0.0106052\ldots$ awrt **0.0106**	A1	1.1b
		(3)	
		(8 marks)	

	Notes
(a)	1st M1 for selecting a Poisson model – written or used. May be implied by 2nd M1 or a correct answer. 2nd M1 for the correct Poisson $\text{Po}(\frac{20}{3})$ or Po(6.67) or better seen and writing or using $1 - \text{P}(C \leqslant 4)$ A1 for awrt 0.794 (correct ans with no incorrect working scores 3/3)
(b)	M1 for selecting a correct model B(4, 0.189) or better (calc: 0.188875…) A1 for using the model to get awrt 0.0219 (correct ans with no incorrect working scores 2/2)
(c)	1st M1 for a correct prob of 1 call (expressions in e or values) (allow 0.31479… or awrt 0.315 or 0.033689… or awrt 0.0337) 2nd M1 for a correct probability statement or expression. E.g. $\text{P}(S = 1 \mid S \sim \text{Po}(\frac{5}{3})) \times \text{P}(T = 1 \mid T \sim \text{Po}(5))$
SC	e.g. $F \sim \text{Po}(\lambda)$ used in (b) to find $\text{P}(F = 0)$ Then if we see $Y \sim \text{Po}(3\lambda)$ and statement $\text{P}(F = 1) \times \text{P}(Y = 1)$ award M0M1 A1 for awrt 0.0106 (correct ans with no incorrect working scores 3/3)

Examiner Comments

A special case was available in part (c) for those students who had the wrong rate for the 5 minute break in part (b) and used this rate in their Poisson distributions in part (c).

Student Response A

4 hours = 80 calls

$\div 12$ $\div 12$

a) X = calls received in 20 mins   20 mins $= \frac{20}{3}$

$X \sim Po\left(\frac{20}{3}\right)$

$P(X > 4) = 1 - P(X \leq 4) = 0.7944$

$1 - 0.2056$

b) Y = no. calls in 5 mins

$Y \sim Po(~~4~~ \frac{4}{3})$. $P(Y=0) = 0.263597$.

$Z \sim B(4, 0.2636)$

$P(Z=3) = 0.05395$

c) ~~P(Z~~ $P(Y=1) = 0.35146$

M = no. calls in 15 mins

$M \sim Po(4)$

$P(M=1) = 0.07326$

$0.35146 \times 0.07326 = 0.02575$.

4/8

Examiner Comments

In part (a) this is correct (M1M1A1).

In part (b) they have an incorrect model for the number of calls in a 5 minute break and so their binomial model is incorrect (M0 and therefore A0). This is a case where the special case might provide a mark in part (c).

In part (c) they have used their incorrect rate for 5 minutes to get the rate for 15 minutes but the calculation they make with these two Poisson models is the right one. We therefore award M0 for the incorrect Poisson models but M1 (using the special case) for attempting to multiply two appropriate probabilities but the answer of course is wrong (A0).

Student Response B

(a) $80 \div 12 = \frac{20}{3}$

$X \sim Po\left(\frac{20}{3}\right)$

$P(X > 4) = 1 - P(X \leq 4)$

$= 1 - 0.2056$

$= 0.7944$

(b) $Y \sim Po\left(\frac{5}{3}\right)$

$P(X=0) = 0.1889$

~~0.18~~ $Z \sim B(4, 0.1889)$

$P(X=3) = 0.0219$

(c) $A \sim Po\left(\frac{5}{3}\right)$

$P(X=1) = 0.3148$

$P(X=0) = 0.1889$

$B \sim Po\left(\frac{15}{3}\right)$

$P(X=1) = 0.0337$

$P(X=0) = 0.0067$

$0.3148 \times 0.0067 + 0.1889 \times 0.0337$

$= 0.0085$

6/8

Examiner Comments

In part (a) this is fully correct (M1M1A1).

In part (b) they have used the correct rate on a Poisson to find the parameter of the correct binomial distribution (M1) and have used a sufficiently accurate value to achieve a final answer correct to 3s.f. (A1).

In part (c) the two correct Poisson models are stated and the probability of 1 missed call worked out for at least one of them (1st M1). This student has then misinterpreted the question and effectively found the probability of exactly 1 missed call in just one of the breaks. This was a fairly common error and scored (M0A0).

Student Response C

~~[illegible]~~ a) $X \sim Po(80)$ per 4 hrs (per 240 mins)

$X \sim Po(\frac{20}{3})$ per 20 minutes

$P(X > 4) = P(X \geqslant 5)$

$= 1 - P(X \leqslant 4)$

$= 0.205627\ldots$

$= 0.206$ (3sf)

b) $X \sim Po(\frac{5}{3})$ per 5 minutes

$P(X=0) = 0.188875\ldots$

$X \sim B(4, 0.188875\ldots)$

$P(X=3) = 0.0219$ (3sf)

c) $X \sim Po(\frac{5}{3})$ per 5 minutes

$P(X=1) = 0.31479\ldots$

~~$X \sim B(4,$~~

$Y \sim Po(5)$ per 15 minutes

~~[illegible]~~

$P(Y=1) = 0.03368\ldots$

$\therefore$ ~~[illegible]~~

$P(X=1) \times P(Y=1) = 0.0106$ (3sf)

7/8

Examiner Comments

In part (a) the correct model is given (M1) and the correct probability expression is used (M1) but they forgot to subtract their value from 1 and so the answer is incorrect (A0).

In part (b) this is fully correct (M1A1).

In part (c) this is a very well presented solution. They have clearly defined two Poisson variables *X* and *Y*, worked out the probability of 1 missed call for each and then given a clear expression using their notation to achieve the answer. All the marks are scored and there is good use of mathematical notation.

Exemplar Question 3

back to Contents Page

3. A biased spinner can land on the numbers 1, 2, 3, 4 or 5 with the following probabilities.

Number on spinner	1	2	3	4	5
Probability	0.3	0.1	0.2	0.1	0.3

The spinner will be spun 80 times and the mean of the numbers it lands on will be calculated.
Find an estimate of the probability that this mean will be greater than 3.25

(6)

(Total for Question 3 is 6 marks)

Mean Score 4.89 out of 6

Examiner Comments

This turned out to be the most successfully answered question on the paper with nearly 70% scoring full marks. Even those who did not spot the need to use the central limit theorem could often score the first 3 marks although some still struggled to use the correct formula for the variance with – E(X), instead of – E(X^2), a common error. Use of correct notation was not good with the statements like $X \sim \text{N}\left(3, \sqrt{\frac{2.6}{80}}\right)$ often being used incorrectly as $\bar{X} \sim \text{N}\left(3, \sqrt{\frac{2.6}{80}}^{2}\right)$. Students should be aware that there is an assessment objective in the new specification that requires "correct use of mathematical notation" and correct handling of normal distribution notation may be required in future assessments.

Mark Scheme

Qu	Scheme	Marks	AO
3.	{ Let X = the number when the spinner is spun} $\mu = \underline{\mathbf{3}}$	B1	1.1b
	$\left[\mathrm{E}(X^2)=\right]0.3+4\times0.1+9\times0.2+16\times0.1\ +25\times0.3$ [= 11.6 or $\frac{58}{5}$]	M1	1.1b
	$\sigma^2\left[=11.6-3^2=\right]\ \underline{\mathbf{2.6}}$	A1	1.1b
	$\bar{X} \approx\sim \mathrm{N}\left(\text{"3"},\sqrt{\dfrac{\text{"2.6"}}{80}}^{\,2}\right)$	M1	2.1
		A1ft	1.1b
	$\mathrm{P}(\bar{X}>3.25)=[\mathrm{P}(Z>1.3867...)=]0.0827589...\text{(calc)}$ awrt $\underline{\mathbf{0.0828}}$	A1	3.4
		(6 marks)	

Notes

	B1 for stating or using mean = 3
	1st M1 for using the given model to attempt $\mathrm{E}(X^2)$ with at least 3 correct products seen
	1st A1 for Var(X) = 2.6 or $\sigma=\sqrt{2.6}=1.6124...$ (awrt 1.61)
ALT	**Use of pgf** (B1 when mean = 3 seen) (M1 when correct $\mathrm{G}''(t)$ seen with attempt at $\mathrm{G}''(1)$) $\mathrm{G}(t)=0.3t+0.1t^2+0.2t^3+0.1t^4+0.3t^5$ $\mathrm{G}'(t)=0.3+0.2t\ +0.6t^2+0.4t^3+1.5t^4$ $\mathrm{G}''(t)=0.2+1.2t+1.2t^2+6t^3$ leading to $\mathrm{G}''(1)=8.6$
	2nd M1 for use of CLT – must use $\bar{X}$ and normal <u>or</u> sight of $\mathrm{N}\left(\text{"3"},\sqrt{\dfrac{\text{"2.6"}}{80}}^{\,2}\right)$ with any letter
	2nd A1ft for a correct mean and variance, ft their 3 and their 2.6 This M1A1ft may be implied by sight of correct st. dev. used in a standardisation leading to P(Z > 1.39) Must see correct use of Z NB $\dfrac{2.6}{80}=0.0325$ and $\sqrt{\dfrac{2.6}{80}}=0.18027...$ so allow e.g. N(3, awrt $(0.180)^2$)
	3rd A1 for using the normal model to find probability awrt 0.0828
ALT	**Use of** $\sum X$ (If see clear attempt at P(ΣX > 260) condone P(ΣX > 260.5) then: 2nd M1 for $\Sigma X\sim$ N(…) <u>or</u> any letter ~N(“240”, $\sqrt{\text{"2.6"}\times80}^{\,2}$) 2nd A1ft for mean = “3”$\times$80 = 240 <u>and</u> variance = “2.6”$\times$80 = 208 May see P(ΣX > 260.5) = 0.077597… but it will only score 2nd M1 2nd A1ft and **3rd A0**

Examiner Comments

The focus of this question was to realise that the central limit theorem was required and that this therefore required calculation of the mean and variance of the population distribution of scores on the spinner.

Although the vast majority of students scored the first 3 marks for finding these correctly, if their values for the mean and variance of scores on the spinner were incorrect but used appropriately in the central limit theorem the second A mark would still be allowed as a follow through.

A few students tried to use a continuity correction here which is not appropriate when using the central limit theorem. This usually led to a final answer of 0.0776 and only lost the final mark.

Student Response A

3) $E(x) = 1\times0.3 + 2\times0.1 + 3\times0.2 + 4\times0.1 + 5\times0.3$

$= 3$

$Var(x) = E(x^2) - (E(x))^2$

$E(x^2) = 1\times0.3 + 2^2\times0.1 + 3^2\times0.2 + 4^2\times0.1 + 5^2\times0.3$

$E(x^2) = 11.6$

$Var(x) = 11.6 - 3^2 = 2.6$

$\sigma = \sqrt{2.6}$

$x \sim N\left(3, \frac{\sqrt{2.6}}{\sqrt{80}}\right)$

$P(x > 3.25) = 1 - P(x \leq 3.25)$

$= 1 - P\left(x \leq \frac{0.25}{\frac{\sqrt{2.6}}{\sqrt{80}}}\right.$

$= 1 - P\left(x \leq \frac{5\sqrt{13}}{13}\right.$

$= 1 - 0.91724$

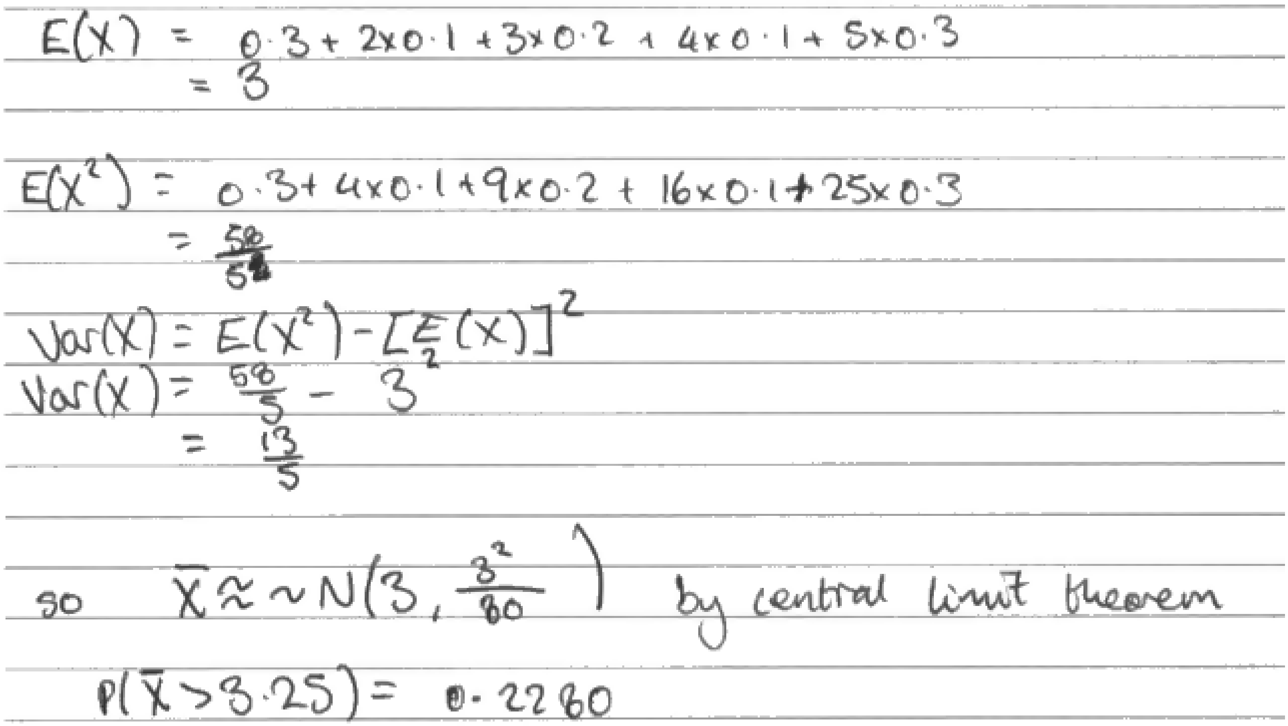
$P(x > 3.25) = 0.08274$

3/6

Examiner Comments

The mean and variance are correct (B1M1A1) and they have attempted to use the central limit theorem (CLT) but not with $\bar{X}$ and they have not stated the normal distribution correctly (standard deviation not variance) and have not used it correctly in the standardisation (we do not see P(Z > 1.39)) so no further marks are possible.

Student Response B



$E(X) = 0{\cdot}3 + 2\times0{\cdot}1 + 3\times0{\cdot}2 + 4\times0{\cdot}1 + 5\times0{\cdot}3$

$= 3$

$E(X^2) = 0{\cdot}3 + 4\times0{\cdot}1 + 9\times0{\cdot}2 + 16\times0{\cdot}1 + 25\times0{\cdot}3$

$= \frac{58}{5}$

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

$\mathrm{Var}(X) = \frac{58}{5} - 3^2$

$= \frac{13}{5}$

so $\bar{X} \approx \sim N\left(3, \frac{3^2}{80}\right)$ by central limit theorem

$P(\bar{X} > 3{\cdot}25) = 0{\cdot}2280$

4/6

Examiner Comments

The mean and variance are correct (B1M1A1) and they have attempted to use the CLT with $\bar{X}$ notation and a normal distribution (M1) but the variance is incorrect (2nd A0) and therefore the answer is incorrect (3rd A0).

Student Response C

$E(x) = 1 \times 0.3 + 2 \times 0.1 + 3 \times 0.2 + 4 \times 0.1 + 5 \times 0.3 = 3$

$G_x(t) = 0.3t + 0.1t^2 + 0.2t^3 + 0.1t^4 + 0.3t^5$

$G_x'(t) = 0.3 + 0.2t + 0.6t^2 + 0.4t^3 + 1.5t^4$

$G_x''(t) = 0.2 + 1.2t + 1.2t^2 + 6t^3$

$Var(x) = G_x''(1) + 3 - 3^2$

$= (0.2 + 1.2 + 1.2 + 6) - 6$

$= 2.6$

$\bar{X}$: mean number the spinner lands on

$X \sim N(3, 2.6)$

$\bar{X} \sim N(3, \frac{2.6}{80})$

$P(\bar{X} > 3.25) = 0.08276$

6/6

Examiner Comments

This fully correct solution has two interesting features.

Firstly the student has used a probability generating function approach to find the mean and variance. Whilst this is not quite as efficient as the standard approach in this situation, it is, of course, perfectly acceptable and an alternative marks scheme was provided.

Secondly, whilst the random variable X (which represents the number the spinner lands on) does not have a N(3, 2.6) distribution, the student uses the $\bar{X}$ notation correctly which we would want to encourage.

Exemplar Question 4

back to Contents Page

4. Liam and Simone are studying the distribution of oak trees in some woodland. They divided the woodland into 80 equal squares and recorded the number of oak trees in each square. The results are summarised in Table 1 below.

Number of oak trees in a square	0	1	2	3	4	5	6	7 or more
Frequency	1	4	21	23	13	11	7	0

Table 1

Liam believes that the oak trees were deliberately planted, with 6 oak trees per square and that a constant proportion p of the oak trees survived.

(*a*) Suggest the model Liam should use to describe the number of oak trees per square.

(2)

Liam decides to test whether or not his model is suitable and calculates the expected frequencies given in Table 2.

Number of oak trees in a square	0 or 1	2	3	4	5	6
Expected frequency	5.53	14.89	24.26	22.24	10.87	2.21

Table 2

(*b*) Showing your working clearly, complete the test using a 5% level of significance. You should state your critical value and conclusion clearly.

(7)

Simone believes that a Poisson distribution could be used to model the number of oak trees per square. She calculates the expected frequencies given in Table 3.

Number of oak trees in a square	0 or 1	2	3	4	5	6 or more
Expected frequency	12.69	16.07	s	14.58	t	9.37

Table 3

(*c*) Find the value of s and the value of t, giving your answers to 2 decimal places.

(4)

(*d*) Write down hypotheses to test the suitability of Simone's model.

(1)

The test statistic for this test is 8.749

(*e*) Complete the test. Use a 5% level of significance and state your critical value and conclusion clearly.

(3)

(*f*) Using the results of these tests, explain whether the origin of this woodland is likely to be cultivated or wild.

(2)

(Total for Question 4 is 19 marks)

Mean Score 12.73 out of 19

Examiner Comments

Part (a) was answered well with the majority identifying a suitable binomial distribution although a few did think a Poisson was appropriate here.

In part (b) the most common error was a failure to pool the last two classes and this often led to the test being carried out using 4 degrees of freedom and the subsequent acceptance of the binomial model.

Part (c) though was answered very well with most students scoring full marks here. There was some impressive, but totally unnecessary, work here to establish the value of λ used by Simone; a number used the given expected frequencies for 2 and 4 oak trees and solved the resulting equations instead of simply using the mean value from the given data.

In part (d) most students incorrectly quoted the value of the parameter; this was penalised in part (d) but condoned for the final mark of part (e).

Part (e) was answered well but some used an incorrect value for the degrees of freedom. There were some good answers to the final part with many appreciating that the suitability of the Poisson model suggested that the oak trees were randomly spread in the woodland and this would imply that it was wild and not cultivated. Whilst many thought that the woodland was probably wild some were unable to give a suitable reason based on the results of the tests.

Mark Scheme

<table>
<tr><th>Qu</th><th>Scheme</th><th>Marks</th><th>AO</th></tr>
<tr><td>4(a)</td><td>[T = no. of oak trees in a square] $T \sim$ Binomial</td><td>M1</td><td>3.3</td></tr>
<tr><td></td><td>$T \sim B(6, p)$</td><td>A1</td><td>1.1b</td></tr>
<tr><td></td><td></td><td>(2)</td><td></td></tr>
<tr><td>(b)</td><td>Expected frequency for 6 is less than 5 so pool: new E_i = 13.08</td><td>M1</td><td>2.1</td></tr>
<tr><td></td><td><table>
<tr><td>$\frac{(O_i - E_i)^2}{E_i}$</td><td>0.051</td><td>2.51</td><td>0.0654</td><td>3.84</td><td>1.85</td></tr>
<tr><td>$\frac{O_i^2}{E_i}$</td><td>4.521</td><td>29.617</td><td>21.805</td><td>7.599</td><td>24.771</td></tr>
</table>
$\sum \frac{(O_i - E_i)^2}{E_i} = 8.313$</td><td>M1,A1</td><td>1.1b x2</td></tr>
<tr><td></td><td>p needed estimating ($\hat{p} = 0.55$) so $\nu = 5 - 2 = 3$; cv 7.815</td><td>B1,B1ft</td><td>1.1b x2</td></tr>
<tr><td></td><td>Significant result, so Liam’s model is not suitable</td><td>M1,A1</td><td>1.1b2.2b</td></tr>
<tr><td></td><td></td><td>(7)</td><td></td></tr>
<tr><td>(c)</td><td>[R = no. of oak trees in a square for Simone’s model] $R \sim$ Po(3.3)</td><td>M1</td><td>3.3</td></tr>
<tr><td></td><td>Correct expression for s or t using Poisson</td><td>M1</td><td>3.4</td></tr>
<tr><td></td><td>s = 17.67 and t = 9.62</td><td>A1,A1</td><td>1.1b x2</td></tr>
<tr><td></td><td></td><td>(4)</td><td></td></tr>
<tr><td>(d)</td><td>H_0: Poisson is a good fit (for no. of oak trees per square)
H_1: Poisson is not a good fit (for no. of oak trees per square)</td><td>B1</td><td>2.5</td></tr>
<tr><td></td><td></td><td>(1)</td><td></td></tr>
<tr><td>(e)</td><td>No pooling needed so degrees of freedom is 6 – 2 = 4</td><td>B1</td><td>1.1b</td></tr>
<tr><td></td><td>Critical value is 9.488 (accept 9.49)</td><td>B1</td><td>1.1a</td></tr>
<tr><td></td><td>Not significant so Poisson (or Simone’s) model is suitable</td><td>B1</td><td>2.2b</td></tr>
<tr><td></td><td></td><td>(3)</td><td></td></tr>
<tr><td>(f)</td><td>Poisson model has better fit so suggests that oak trees occur at random
Or binomial suggests deliberately planted or cultivated</td><td>B1</td><td>2.2b</td></tr>
<tr><td></td><td>Therefore the forest is likely to be wild not cultivated</td><td>B1</td><td>3.5a</td></tr>
<tr><td></td><td></td><td>(2)</td><td></td></tr>
<tr><td></td><td></td><td colspan="2">(19 marks)</td></tr>
</table>

	Notes
(a)	M1 for choosing binomial A1 for B(6, *p*) can be in words and allow B(6, 0.55)
(b)	1st M1 for pooling last 2 classes (E_i = 13.08 but accept 13.1) 2nd M1 for at least 3 correct values or expressions. Either row to at least 2 s.f. 1st A1 for awrt 8.31 (8.31 gets 3/3) [NB no pooling gives awrt 16.8458.. and implies M0M1A0] 1st B1 for 3 degrees of freedom 2nd B1ft for critical value of 7.815 (e.g. $\nu = 4$ use 9.488) 3rd M1 for a correct conclusion (non-contextual ignore any contradictory contextual comments for this mark) based on their cv and their test statistic This mark can be implied by a fully correct solution ending with correct contextual conclusion 2nd A1 for correct conclusion in context with **all other marks scored**
(c)	1st M1 for selecting a correct model Po(3.3) [Allow Po(awrt 3.3)] 2nd M1 for use of the model with an expression or correct value for *s* or *t* 1st A1 for one correct 2nd A1 for both correct (allow awrt 2dp)
(d)	B1 for correct hypotheses must mention Poisson: use of Po(3.3) is B0
(e)	1st B1 for correct degrees of freedom $\nu = 4$ only 2nd B1 for selecting correct critical value (9.488 only) 3rd B1 for not significant conclusion based on 8.749 vs their cv (condone use of Po(3.3) here)
(f)	1st B1 for choosing Poisson as better or stating Poisson implies wild or bino'l implies cultivated 2nd B1 (dep on rejecting bin and accepting Poisson) for clearly stating woodland is wild If the tests give the same results then 2nd B0 automatically

Examiner Comments

The final A mark in part (b) was dependent upon all the other marks being scored. We often adopt this practice for the final mark in hypothesis tests.
The final mark in part (f) was dependent upon the student's tests leading to a rejection of a binomial model and "acceptance" of the Poisson model. A number of students, due to errors in previous parts, failed to reject either model and were not therefore in a position to make a clear statement about the likely origin of the woodland although they could still achieve the first mark for a suitable interpretation.

Student Response A

a) Exponential

b) H_0: model is suitable

H_1: model is not suitable

$X^2_{CALC} = \sum \frac{O_i^2}{E_i} - n$

$= \frac{5^2}{5.53} + \frac{21^2}{14.89} + \frac{23^2}{24.26} + \frac{13^2}{22.24} + \frac{4^2}{10.87} - 80$

$= 16.8455\,5076$

Degrees of freedom $= 7 - 1 - 1 - 1 = 4$

$\chi^2_4(5\%) = 9.488$

$16.846 > 9.488$ therefore there is sufficient evidence to reject H_0, accept H_1: model is not suitable

c) $X \sim Po(\lambda)$

$P(x=3) = e^{-\lambda}\left(\frac{\lambda^3}{3!}\right) \times 80 = s$

$= \frac{40\lambda^3}{3e^{\lambda}} = \frac{40\lambda^2}{e^{\lambda}} \times \frac{\lambda}{3}$

$= 16.07 \times \frac{\lambda}{3}$

$P(x=5) = t = e^{-\lambda}\left(\frac{\lambda^5}{5!}\right) \times 80$

$= \frac{2\lambda^5}{3e^{\lambda}} = \frac{40\lambda^2}{e^{\lambda}} \times \frac{\lambda^3}{60}$

$= 16.07 \times \frac{\lambda^3}{60}$

$P(x=2) = 11.07 = e^{-\lambda}\left(\frac{\lambda^2}{2!}\right) \times 80$

$16.07 = \frac{40\lambda^2}{e^{\lambda}}$ $\quad e^{\lambda} = \frac{40\lambda^2}{16.07}$

$P(x=4) = 14.88 = e^{-\lambda}\left(\frac{\lambda^4}{4!}\right) \times 80$

$14.58 = \frac{10\lambda^4}{3e^{\lambda}}$

$14.58 = \frac{10\lambda^4}{\frac{120\lambda^2}{16.07}}$ $\quad 14.88 = \frac{16.07}{200}\lambda^2$

$\lambda = 3.1490\,21153$

c) $P(X \geq 6) = 1 - P(X \leq 5)$

~~9.37~~

~~$\frac{9.37}{80} = 1 - (\frac{t}{80}$~~

$9.37 = 80 - (t + s + 14.38 + 16.07 + 12.64)$

$9.37 = 80 - t - s - 43.34$

$t + s = 27.29$

$16.07\left(\frac{\lambda}{3}\right) + 16.07\left(\frac{\lambda^3}{60}\right) = 27.29$

$321.4\lambda + 16.07\lambda^3 = 1637.4$

$16.07\lambda^3 + 321.4\lambda - 1637.4 = 0$

$\lambda = 3.29914023 9$

$t = \frac{16.07 \times (3.299...)^3}{60}$ $\quad$ $s = 16.07 \times \frac{3.299...}{3}$

$t = 9.61760 5453$ $\quad$ $s = 17.67239455$

~~s = 17.6 (2dp)~~ $\quad$ $\underline{t = 9.62 \text{ (2dp)}}$

$\underline{s = 17.67 \text{ (2dp)}}$

d) H_0: can be modelled using $X \sim Po(\lambda)$

H_1: can not be ~~be~~ modelled using $X \sim Po(\lambda)$

$X^2_{calc} = \sum \frac{O_i^2}{E_i} - n$

$= \frac{5^2}{17.64} + \frac{21^2}{16.07} + \frac{23^2}{17.67} + \frac{13^2}{14.38} + \frac{11^2}{9.62} + \frac{2^2}{9.37} - 80$

$= 8.748881222$

$8.748.. < 8.749$ ∴ sufficient evidence to reject ~~accept~~ H_0, ~~Not a suitable model~~ accept H_1: it is not a suitable model

e) Wild as it is much closer to the test statistic found

8/19

Examiner Comments

In part (a) an incorrect model is chosen (M0A0).

In part (b) they have not pooled the last two groups (M0) but do have at least 3 correct expressions for the test statistic using the $\sum \frac{O_i^2}{E_i} - N$ formula (M1) but the test statistic is incorrect (A0). Their degrees of freedom are not equal to 3 (B0) but they do give the correct critical value for their degrees of freedom (B1ft). They have made a correct non-contextual conclusion (M1) and have said that the model is not suitable, but the final A mark is only available if all the other marks are scored (A0). This was a fairly common solution and this failure to carry out the pooling meant a penalty of 4 marks.

In part (c) this student was one who did not realise that Simone's expected frequencies would have the same mean as the original data and "reversed engineered" the problem to find the value of λ. There is some impressive, but unnecessary, work here that happily leads to the correct value of λ and they then go on to complete part (c) correctly (M1M1A1A1).

In part (d) suitable hypotheses were given (B1).

In part (e) they do not state the degrees of freedom as 4 (B0) and do not have the correct critical value of 9.488 (B0). They seem to be treating the given value of 8.749 as the critical value and have attempted to calculate their own test statistic. So, there is no evidence of comparing a critical value with 8.749 (B0).

In part (f) since they have rejected both models only the first mark is possible. This requires a statement that the Poisson model suggests woodland is wild or binomial suggests it is cultivated. They do not have either statement (B0B0).

Student Response B

~~B)~~ 4) a) X = number of oak trees per square

$X \sim B(80, 6p)$

b) n° of oak trees / square 0 1 2 3 4 5 6 7≥

Observed frequency $\underbrace{1 \quad 4}_{5}$ 21 23 13 $\underbrace{11 \quad 7}_{18}$ 0

Expected frequency ~~$\underbrace{5.83 \; 14.89}$ 24.26 22.24 10.87~~

5.53 14.89 24.26 22.24 $\underbrace{10.87 \quad 2.21}_{\text{combine}}$

13.08

$$\cancel{\Sigma}\,\frac{(O_i - E_i)^2}{E_i}$$

0.0508 (0–1) 3.839 (4)

2.507 (2) 1.8506 (5≥)

0.0654 (3)

$$\Sigma \frac{(O_i - E_i)^2}{E_i} = X^2 = 8.31295 \approx 8.31 \text{ (3s.f.)}$$

$\cancel{X}^2$ v ~~n~~ $= 5 - 1 = 4$ ~~X~~ $X^2_4 (5\%) = 9.488$

(critical ~~co~~ value).

H_0: Liam's model is suitable (expected and observed data are the same)

H_1: Liam's model is not suitable (expected and observed data are different)

$8.31 < 9.488$ Accept H_0 * Liam's model is suitable for this data.

* there is sufficient evidence to suggest that...

c) $P(X = x) = e^{-\lambda} \dfrac{\lambda^x}{x!}$ $P(X = 2) = e^{-\lambda} \dfrac{\lambda^2}{2} = 16.07$

$\Rightarrow e^{-\lambda} \lambda^2 = 32.14$

$P(X = 4) = e^{-\lambda} \dfrac{\lambda^4}{24} = 14.58 \rightarrow e^{-\lambda} \lambda^4 = 349.92$

$e^{-\lambda} = \frac{32.14}{\lambda^2}$ $\quad e^{-\lambda} = \frac{349.92}{\lambda^4}$

$\frac{32.14}{\lambda^2} = \frac{349.92}{\lambda^4} \rightarrow 32.14\lambda^2 = 349.92$

$\lambda^2 = \frac{349.92}{32.14}$,

$\lambda = \sqrt{\frac{349.92}{32.14}}$

~~Pt~~ $X \sim Po\left(\sqrt{\frac{349.92}{32.14}}\right)$ $\quad P(X=3) = 0.221$ (3s.f.)

$P(X=5) = 0.120$ (3s.f.)

s = 0.221 × 80 = ~~17.68~~ 17.67 (~~3~~ 2d.p.)

t = 0.120 × 80 = 9.62 (2d.p.)

d) H0: this model is suitable for the data i.e observed and ~~H1: this model is not suitable for the data~~ expected data are the same

H1: this model is not suitable i.e. expected and observed data are different.

e) $X^2 = 8.749$ ~~?~~ $\nu = 6-1=5$

$\chi^2_5(5\%) = 11.070 > X^2$

Accept H0, there is sufficient evidence to suggest that Jiménez's model is suitable.

f) Both of these models were suitable for the data and both models ~~used~~ assumed that the trees were deliberately planted ~~and that~~ (6 trees per square) and that a proportion p survived. Therefore the woodland was probably cultivated.

11/19

Examiner Comments

In part (a) a binomial model is given (M1) but the value of n and the probability are both incorrect (A0).

In part (b) they have pooled the last two groups (M1) and have a correct test statistic (M1A1). The degrees of freedom is incorrect (B0) but they have a correct follow through (ft) value for the critical value (B1ft) and have made a correct statement using their test statistic and their critical value (M1) but the final A mark is only available if all the other marks are scored (A0)

In part (c) a slightly different "reverse engineering" approach is used here which once again leads to a correct value for λ and a fully correct solution to part (c) (M1M1A1A1).

In part (d) their hypotheses do not clearly state that a Poisson distribution is being proposed as a suitable model for the number of oak trees per square. There needed to be some explicit mention of the Poisson distribution (B0).

In part (e) the degrees of freedom and critical value were both incorrect (B0B0) [There was no follow through mark offered here]. There is a correct comparison of the given value of 8.749 with their critical value and this leads to a not significant result and so the final mark was awarded (B1).

In part (f) since they have not rejected either model the final mark is lost (B0) and since they have said that both models suggest the woodland is cultivated they cannot score the 1st B mark either (B0). A correct interpretation of a binomial model suggesting cultivated would have scored the mark.

Student Response C

a) $X \sim B(6, p)$ $E(X) = 3.3$ $p = 0.55$

Binomial distribution $B(6, 0.55)$.

b) H_0: Binomial distribution $n=6$, $p=0.55$ is a good model

H_1: Binomial distribution $n=6$, $p=0.55$ is a bad model.

No. of Oak trees	0 or 1	2	3	4	5≤
Expected	5.53	14.89	24.26	22.24	13.08
Observed	5	21	23	13	18.

$\nu = 5 - 1 - 1$

$= 3.$

$\chi_3^2(5\%) = 7.815.$

Test stat $= \Sigma \frac{O_i^2}{E_i} - N$

$= 88.3129925\ldots - 80$

$= 8.3129925\ldots$

8221 $8.3129.. > 7.8145$

$\therefore$ there is enough evidence to reject H_0.

Binomial $n=6$, $p=0.55$ is a bad model.

c) $X \sim Po(np)$

$X \sim Po(3.3)$

$P(X=3) = \frac{S}{80} = 0.22091173...$

$P(X=5) = \frac{t}{80} = 0.12028643...$

$S = 17.6729...$

$= 17.67$ (2dp)

$t = 9.6229...$

$= 9.62$ (2dp)

d) H_0: $Po(3.3)$ is a good model

H_1: $Po(3.3)$ is a bad model

e) $\nu = 6 - 2$

$= 4.$

$\chi^2_4 (5\%) = 9.488$

$8.744 < 9.488.$

$\therefore$ not enough evidence to reject H_0.

$Po(3.3)$ is a good model.

f) Woodlands is likely to be wild since it is modelled by a poisson. $\therefore$ each event is random with a fixed mean.

18/19

Examiner Comments

In part (a) we would have accepted either B(6, *p*) or B(6, 0.55) here (M1A1).

In part (b) this is a fully correct solution. We did not require the hypotheses here though of course strictly the value of the parameter should not be included. The question specifically said that working should be shown and critical value and conclusion clearly stated. They have shown the intention of pooling, and the correct test statistic without details of its calculation and that was acceptable. The critical value and conclusion are clearly stated and correct as well so all 7 marks were scored.

In part (c) the correct Poisson model is stated (with no unnecessary calculations) and the correct values for *s* and *t* are given (M1M1A1A1).

In part (d) the hypotheses mention a Poisson distribution but the parameter of 3.3 is included. Since this value was calculated from the observed data, and the degrees of freedom reduced accordingly, it should not be included in null hypothesis and so this mark was lost (B0).

In part (e) the degrees of freedom, critical value and conclusion (non- rejection of the Poisson model) are all correct (B1B1B1). Since we have penalised mentioning the value of the parameter as 3.3 in the hypotheses in part (d) we would always condone it in the conclusion.

In part (f) they have realised that a Poisson model suggests that the trees are distributed randomly (B1) and have therefore concluded that the woodland is wild. The binomial model was rejected in part (b) and a Poisson "accepted" in part (e) so the final mark is scored too (B1).

Exemplar Question 5

back to Contents Page

5. Information was collected about accidents on the *Seapron* bypass. It was found that the number of accidents per month could be modelled by a Poisson distribution with mean 2.5

Following some work on the bypass, the numbers of accidents during a series of 3-month periods were recorded. The data were used to test whether or not there was a change in the mean number of accidents per month.

(*a*) Stating your hypotheses clearly and using a 5% level of significance, find the critical region for this test. You should state the probability in each tail.

(5)

(*b*) State P(Type I error) using this test.

(1)

Data from the series of 3-month periods are recorded for 2 years.

(*c*) Find the probability that at least 2 of these 3-month periods give a significant result.

(3)

Given that the number of accidents per month on the bypass, after the work is completed, is actually 2.1 per month,

(*d*) find P(Type II error) for the test in part (*a*).

(3)

(Total for Question 5 is 12 marks)

Mean Score 9.29 out of 12

Examiner Comments

Part (a) was usually answered well but a few students were going from P($X \leq 13$) to $X \geq 13$ as the critical region and occasionally students were using Po(2.5) instead of Po(7.5).

Most students knew how to find the probability in part (b).

Many identified the binomial model in part (c) but some were confused about the appropriate value for n (3 being a common error) and others made rounding errors which meant that their final answer was not accurate enough.

In the final part most knew what a Type II error was and how to find its probability but sometimes a Po(2.1) model was used rather than the required Po(6.3).

Mark Scheme

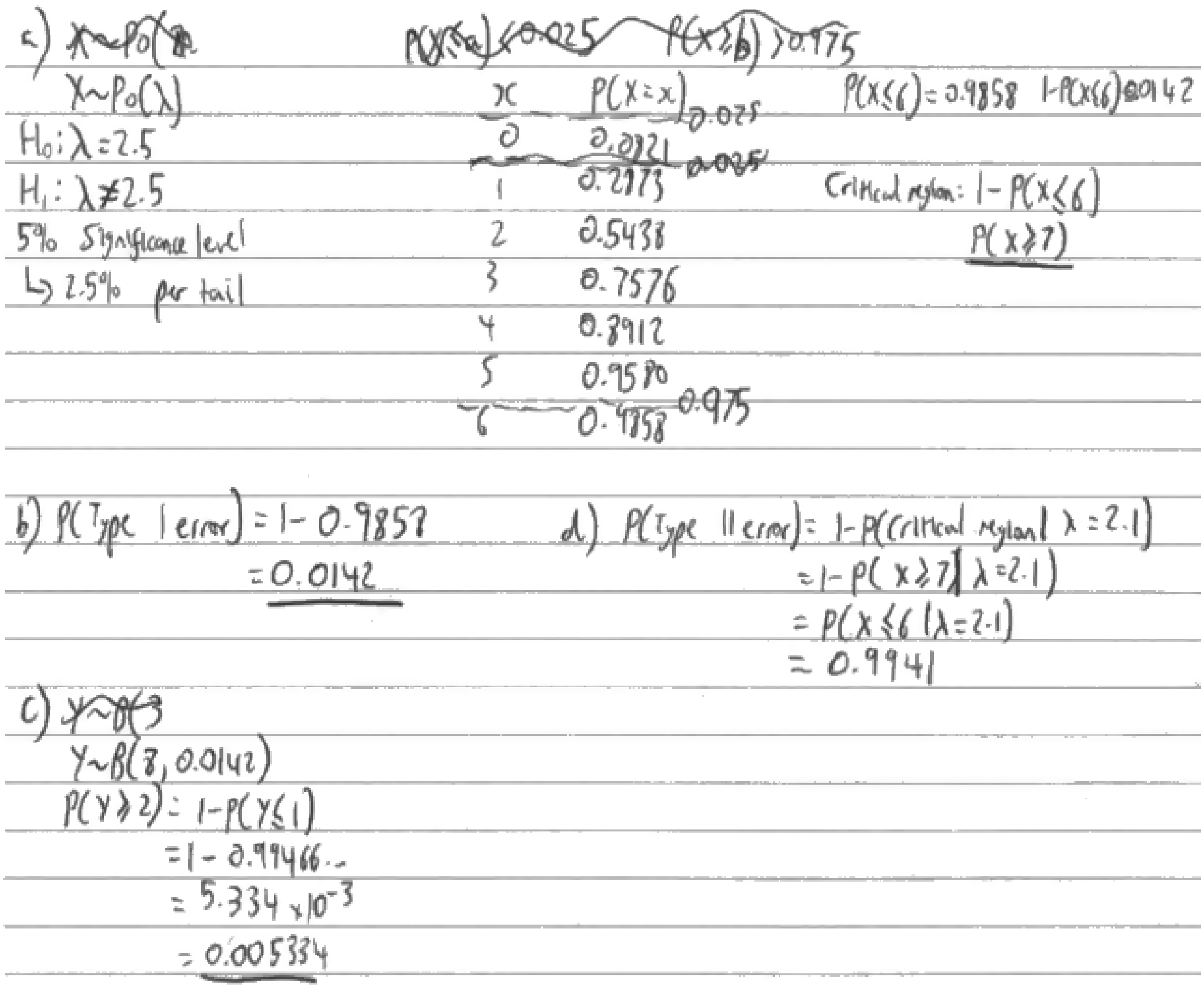
Qu	Scheme	Marks	AO
5(a)	$H_0 : \lambda = 2.5$ (or $\mu = 7.5$) $H_1 : \lambda \neq 2.5$ (or $\mu \neq 7.5$)	B1	2.5
	[X = no. of accidents in a 3-month period] $X \sim \text{Po}(7.5)$	M1	3.3
	$\text{P}(X \leq 2) = 0.0203$ (calc: 0.020256…) { or $\text{P}(X \leq 3) = 0.0591$}		
	$\text{P}(X \leq 13) = 0.9784$ so $\text{P}(X \geq 14) = 0.0216$ (calc: 0.0215646…) {or $\text{P}(X \geq 15) = 0.0103$}	M1	3.4
	Giving Critical region of: $\boldsymbol{X \leq 2}$	A1	1.1b
	$\boldsymbol{X \geq 14}$	A1	1.1b
		(5)	
(b)	[0.0203 +0.0216] =awrt **0.0419** or (calc: 0.041821366… awrt **0.0418**)	B1ft	1.2
		(1)	
(c)	[Let M = no of 3-month periods with a significant result]		
	$M \sim \text{B}(8, \text{“}0.0419\text{”})$	M1	3.3
	$[\text{P}(M \geq 2)] = 1 - \text{P}(M \leq 1)$	M1	1.1b
	$[= 1 - 0.9584…]$		
	=0.04153…(calc: 0.041394…) [**0.04139~ 0.04154]**	A1cso	1.1b
		(3)	
(d)	$Y \sim \text{Po}(6.3)$	M1	3.3
	P(Type II error) = $\text{P}(3 \leq Y \leq 13)$ or $\text{P}(Y \leq 13) - \text{P}(Y \leq 2)$	M1	3.4
	[= 0.9945147… – 0.049846…]		
	= 0.9446… awrt **0.945**	A1	1.1b
		(3)	
		(12 marks)	

	Notes
(a)	B1 for both hypotheses in terms of λ or μ (either way around) 1st M1 for selecting the correct Po model. Sight or use of Po(7.5) may be implied by 2nd M1 2nd M1 for using the correct model to find one of these probs with correct label (2s.f. or better) 1st A1 for one end correct 2nd A1 for a fully correct CR Allow any letter, even CR ≤ 2 or set notation but **not** $\text{P}(X \leq 2)$ Can have $X < 3$ and $X > 13$ etc
(b)	B1ft for awrt 0.0419 or awrt 0.0418 or ft addition of their **two** probs provided both are $0 < \text{prob} < 0.025$ (awrt 3s.f.)
(c)	1st M1 for selecting a correct binomial model, ft their answer to part (b) 2nd M1 for a correct probability statement of $1 - \text{P}(M \leq 1)$ **dep on** a binomial selected A1cso for answer in range [0.04139, 0.04154] **dep on** use of B(8, “0.0419”) or better
(d)	1st M1 for selecting a Po(6.3) model 2nd M1 for a correct probability statement using their Poisson model and their CR in (a) which may have just one tail. A1 for awrt 0.945

Examiner Comments

In part (b) the mark could be awarded as a follow through from part (a) but this was dependent on the addition of two probabilities (so those with a one tailed test could be awarded it) and also on the probabilities both being between 0 and 0.025 (since the critical values were for a two-tailed test at 5% and the probability in each tail should therefore be less than 2.5% since the question did not specify any alternative approach such as “as close as possible to 0.025”).

Student Response A



c) ~~$X \sim Po($~~

$X \sim Po(\lambda)$

$H_0 : \lambda = 2.5$

$H_1 : \lambda \neq 2.5$

5% Significance level

↳ 2.5% per tail

$P(X \leq a) \leq 0.025$ $P(X \geq b) \geq 0.975$

x	$P(X=x)$
0	0.0821
1	0.2873
2	0.5438
3	0.7576
4	0.8912
5	0.9580
6	0.9858

0.025

0.025

0.975

$P(X \leq 6) = 0.9858$ $1 - P(X \leq 6) = 0.0142$

Critical region: $1 - P(X \leq 6)$

$\underline{P(X \geq 7)}$

b) $P(\text{Type I error}) = 1 - 0.9858$

$= \underline{0.0142}$

d) $P(\text{Type II error}) = 1 - P(\text{Critical region} \mid \lambda = 2.1)$

$= 1 - P(X \geq 7 \mid \lambda = 2.1)$

$= P(X \leq 6 \mid \lambda = 2.1)$

$= 0.9941$

c) ~~$Y \sim B(3$~~

$Y \sim B(8, 0.0142)$

$P(Y \geq 2) = 1 - P(Y \leq 1)$

$= 1 - 0.99466...$

$= 5.334 \times 10^{-3}$

$= \underline{0.005334}$

4/12

Examiner Comments

In part (a) their hypotheses are correct (B1) but they have not stated (or used) a correct model to work out the probabilities. In fact Po(2.5) seems to have been used as their statement that $P(X \leq 6) = 0.9858$ confirms. No other marks were available in part (a).

In part (b) their use of Po(2.5) meant that their critical region effectively only had one tail and so they were not adding two probabilities together here (B0). The question in part (a) did explicitly state that there should be two tails and two probabilities so there was a pointer at this stage that an error had been made.

In part (c) they have used their incorrect probability from part (b) in a correct binomial model (M1) and attempt to find a correct probability (M1) but the answer, of course, is incorrect (A0).

In part (d) they are using Po(2.1) instead of Po(6.3) (M0) but they are attempting to find an appropriate probability using their critical region from part (a) (M1) but the answer is incorrect.

Student Response B

5) a) $H_0: \mu = 2.5$ ~~X = number of accidents in 3 months~~

$H_1: \mu \neq 2.5$ ~~$X \sim Po(2.5 \times 3)$~~ ~~month~~

~~$X \sim Po(7.5)$~~

X = number of accidents in 3 months

$X \sim Po(7.5)$ ~~$P(X \geq 6) = 0.378$~~

~~$P(X \geq 5) < 0.244 \leq 0.$~~

$P(X \leq 3) = 0.0591$ $P(X \geq 12) = 0.957$

$P(X \leq 2) = 0.0203 < 0.025$ $P(X \geq 13) = 0.978 > 0.975$

Critical region: ~~$X \geq 3$ and $X \leq 13$.~~

$X \leq 2$ and $X \geq 13$.

b) P(type I error) $= 0.0203 + (1 - 0.978)$

$= 0.0423$

c) $X \sim B(8, 0.0423)$ $P(X \geq 2) = 1 - P(X \leq 1)$

$= 1 - 0.9577$

$= 0.0423$

d) type II error = accepting H_0 when it is false.

$X \sim Po(2.1 \times 3)$

$X \sim Po(6.3)$ ~~$P(X \geq 4) = 1 - P(X \leq 3)$~~

~~$= 1 - 0.12637$~~

~~$= 0.87363$~~

$P(X \leq 12) = 0.98725$

$P(X \leq 2) = 0.049846 \ldots$

$P(X \geq 13) = 1 - P(X \leq 12)$

$= 1 - 0.98725 = 0.01275$

P(type II error) $= 1 - 0.049846 - 0.01275$

$= 0.937$ (3s.f.)

9/12

Examiner Comments

In part (a) the hypotheses are correct (B1) and the correct model is chosen (M1). This is used correctly to find at least one correct probability (M1) and the lower tail of the critical region is correct (A1) but the upper tail is not correct (A0).

In part (b) this is a correct follow through using their probabilities from part (a) (B1ft).

In part (c) the model is a correct follow through (M1) and they have a correct probability statement (M1) but the answer is not in the required range (A0).

In part (d) the correct model is stated (M1) and used correctly to find the probability of a Type II error from their CR in part (a) (M1) but the answer is incorrect (A0).

Student Response C

a) $X \sim Po(7.5)$ accidents per month

$H_0: \lambda = 7.5$

$H_1: \lambda \neq 7.5$

Assume H_0:

~~px~~ $p(X \leq c_1) < 0.025$

$p(X \leq 2) = \underline{0.0203}$

$c_1 = 2$

$p(X \geq c_2) < 0.025$

$p(X \geq 14) = 1 - p(X \leq 13) = \underline{0.0217}$

$c_2 = 14$

Critical region: $2 \leq X$ or $X \geq 14$

b) $p(\text{type I error}) = 0.0203 + 0.0217$

$= p(\text{reject } H_0 / H_0 \text{ true}) = \underline{0.0418}$

c) $p(\text{significant result}) = 0.0418$

$Z \sim B(8, 0.0418)$

$P(Z \geq 2) = 1 - p(Z \leq 1)$

$= \underline{0.0414}$

d) $p(\text{type II error}) = p(\text{Accept } H_0 / H_0 \text{ false})$

$=$ ~~$p(X \leq$~~

$p(3 \leq X \leq 13) / Z \sim Po(6.3)$

$p(Z \leq 13) / Z \sim Po(6.3) = \underline{0.9945}$

$p(Z \leq 2) / Z \sim Po(6.3) = 0.0498$

$0.9945 - 0.0498 =$ ~~0.944~~ $\underline{0.945}$

0.945

11/12

Examiner Comments

In part (a) the hypotheses are correct (B1). We allow either λ or μ to be used for the parameter of the Poisson distribution. The correct model is stated and used to find at least one correct probability (M1M1) and the upper tail of the critical region (CR) is correct (A1) but they have stated the lower tail incorrectly: $X \geq 2$ instead of $X \leq 2$ (A0).

In part (b) 0.0418 is an acceptable answer here (B1). They have clearly used the correct critical region and we would usually allow a correct answer for a new part of the question.

In part (c) this is fully correct (M1M1A1).

In part (d) this is fully correct (M1M1A1).

Exemplar Question 6

back to Contents Page

6. The discrete random variable *X* has probability generating function

$$G_X(t) = k\ln\left(\frac{2}{2-t}\right)$$

where *k* is a constant.

(*a*) Find the exact value of *k*.

(1)

(*b*) Find the exact value of Var(*X*).

(7)

(*c*) Find P(*X* = 3).

(4)

(Total for Question 6 is 12 marks)

Mean Score 7.36 out of 12

Examiner Comments

Part (a) was answered very well and most obtained the correct value for *k* though sometimes this was incorrectly expressed as –ln2.

In part (b) it was clear that most students knew how to use the probability generating function to find the variance but there were a number who struggled to differentiate the given function correctly. A significant minority wrote G(*t*) as *k*[ln2 –ln (2 – *t*)] and they fared better although some did not realise that ln2 was a constant and a spurious $\frac{1}{2}$ appeared. Those who did differentiate correctly and simplified their answer invariably were able to complete this part of the question.

Part (c) was more of a challenge but a good number attempted this. Many realised that a Maclaurin expansion was required and finding the third derivative was the most common approach. Unfortunately, some forgot to divide by 3! But nearly a quarter of the students secured full marks on this question.

Mark Scheme

Qu	Scheme	Marks	AO
6 (a)	$\mathrm{G}(1) = 1 \Rightarrow k\ln 2 = 1$ so $k = \underline{\underline{\frac{1}{\ln 2}}}$	B1	2.1
		(1)	
(b)	$\left\{\mathrm{G}(t) = \frac{1}{\ln 2}[\ln 2 - \ln(2-t)]\right\} \Rightarrow \mathrm{G}'(t) = \frac{1}{\ln 2}\left[\frac{1}{2-t}\right]$ or $\frac{1}{\ln 2}(2-t)^{-1}$	M1 A1	2.1 1.1b
	[E(X) =] $\mathrm{G}'(1) = \frac{1}{\ln 2}$	A1	1.1b
	$\mathrm{G}''(t) = \frac{1}{\ln 2} \times \left[\frac{1}{(2-t)^2}\right]$	M1 A1	2.1 1.1b
	$\mathrm{Var}(X) = \mathrm{G}''(1) + \mathrm{G}'(1) - [\mathrm{G}'(1)]^2 = \frac{1}{\ln 2} + \frac{1}{\ln 2} - \left(\frac{1}{\ln 2}\right)^2$	M1	2.1
	$= \underline{\frac{1}{\ln 2}\left(2 - \frac{1}{\ln 2}\right)}$	A1	1.1b
		(7)	
(c)	P(X = 3) = coefficient of t^3 by Maclaurin need $\mathrm{G}'''(0)$	M1	3.1a
	$\mathrm{G}'''(t) = \frac{1}{\ln 2}\frac{2}{(2-t)^3}$	A1ft	1.1b
	$\mathrm{P}(X = 3) = \frac{\mathrm{G}'''(0)}{3!}$	M1	3.2a
	$= \frac{\frac{1}{4\ln 2}}{6} = \frac{1}{24\ln 2} = 0.0601122\ldots$ awrt $\underline{\mathbf{0.0601}}$	A1	1.1b
		(4)	
		(12 marks)	

	Notes
(a)	B1 for finding k (must be exact)
(b)	1st M1 for an attempt to differentiate G(t) e.g. $A(2-t)^{-1}$ (o.e.) 1st A1 for a correct first derivative (condone k or use of $\frac{1}{\ln 2}=$ awrt 1.44) 2nd A1 for correct E(X) or $\text{G}'(1)$ (allow awrt 1.44 calc: 1.442695…but not k) seen anywhere 2nd M1 for attempting second derivative (ft their $\text{G}'(t)$) 3rd A1 for a correct 2nd derivative (condone k or use of $\frac{1}{\ln 2}=$ awrt 1.44) 3rd M1 for a correct method for Var(X) (some substitution into the correct formula) 4th A1 for $\frac{1}{\ln 2}\left(2-\frac{1}{\ln 2}\right)$ o.e. but must simplify i.e. collect like terms [Mark final answer – penalise incorrect log work etc] NB 0.8040211.. is A0 unless exact answer seen
(c)	1st M1 for a suitable strategy to solve the problem (finding link with Maclaurin) Need mention of coefficient of t^3 **and** [$\text{G}'''(t)$ or $\text{G}'''(0)$](condone $\text{G}'''(1)$) 1st A1ft for 3rd derivative , ft their 2nd derivative in (b) (provided $\text{G}''(t)$ not const) Correct $\text{G}'''(t)$ or $\text{G}'''(0)$ scores 1st M1 1st A1ft 2nd M1 for translating Maclaurin to probability (a correct expression) 2nd A1 for $\frac{1}{24\ln 2}$ or awrt 0.0601
ALT	**Log series** 1st M1 attempt to write G(t) in suitable form as far as: $k[\ln 2-\ln(2[1-\frac{t}{2}])]$
	1st A1 reaching $-k\ln(1-\frac{t}{2})$ 2nd M1 use of – ln(1 – x) series (<u>some</u> correct substitution) NB $\text{G}(t)=\frac{1}{\ln 2}\left(\frac{t}{2}+\frac{t^2}{8}+\frac{t^3}{24}+...\right)$

Examiner Comments

Part (c) required students to draw on their knowledge of pure mathematics and use a Maclaurin expansion or the series expansion for ln(1 – x). Whilst a good number identified this was required, accurate differentiation was sometimes a problem.

The first A mark in part (c) was a follow through for attempting to differentiate their second derivative from part (b) provided this second derivative was not a constant.

Student Response A

(a) $G_x'(1) = 1$

$$\therefore \quad k \ln\left(\frac{2}{2-1}\right) = 1$$

$$k = \frac{1}{\ln(2)}$$

(b) $$\text{Var}(X) = G''_x(1) + G'_x(1) - [G'_x(1)]^2$$

$$G_x(t) = \ln\left(\frac{2}{2-t}\right) \times \frac{1}{\ln(2)}$$

$$G'_x(t) = \frac{1}{\ln(2)} \times \frac{1}{\left(\frac{2}{2-t}\right)} \times -2$$

$$\frac{-x(2-t)}{x\ln(2)} = \frac{t-2}{\ln(2)}$$

$$G''_x(t) = \frac{1}{\ln(2)}$$

$$\text{var}(x) = \frac{1}{\ln(2)} + \frac{-1}{\ln(2)} - \frac{1}{2\ln^2}$$

$$\text{Var}(x) = \frac{-1}{2\ln(2)}$$

(c) $P(X=3) =$

3/12

Examiner Comments

In part (a) they have a correct value for k (B1).

In part (b) their attempt at finding $\mathrm{G}'(t)$ does not lead to a function of the form $A\left(2-t\right)^{-1}$ (M0A0A0). Students who did not write $\ln\left(\frac{2}{2-t}\right)$ as $\ln 2-\ln(2-t)$ often ran into difficulties at this stage. They do differentiate their $\mathrm{G}'(t)$ correctly though to obtain their $\mathrm{G}''(t)$ (M1) but it is, of course incorrect (A0). They quote a correct formula using the probability generating function for Var(X) and there is some correct substitution (M1) but their answer is incorrect (A0).

In part (c) there is no attempt at finding a series expansion of G(t) (M0A0M0A0).

Student Response B

a) $G_x(1) = 1$

$k \ln\left(\frac{2}{2-1}\right) = 1$

$\ln 2 = \frac{1}{k}$

$k = \frac{1}{\ln 2}$

~~b) $G'_x(t) = \frac{1}{\ln 2} \frac{d}{dt}\left(\ln\left(\frac{2}{2-t}\right)\right)$~~

~~$u = 2-t$~~ ~~$y = \ln\left(\frac{2}{u}\right)$~~

~~$\frac{du}{dt} = -1$~~ ~~$\frac{dy}{du} = \frac{u}{2} = \frac{2-t}{2}$~~

~~$\frac{dy}{du} = \frac{du}{dt} \times \frac{dy}{du} = \frac{2-t}{2} \times -1 = \frac{t-2}{2}$~~

~~$G'_x(t) = \frac{t-2}{2\ln 2}$~~

~~$G''_x(t) = \frac{1}{2\ln 2}$~~

~~$\text{Var}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2$~~

~~$= \frac{1}{2\ln 2} + \left(\frac{-1}{2\ln 2}\right) - \left(\frac{-1}{2\ln 2}\right)^2$~~

~~$= - \left(\frac{1}{2^2(\ln 2)^2}\right)$~~

~~$=$~~

~~$G'_X(t) = \frac{1}{\ln 2} \times \ln\left(\frac{2}{2-t}\right)$~~

$G'_X(t) = \frac{1}{\ln 2} \times \frac{d}{dt}\left(\ln\left(\frac{2}{2-t}\right)\right.$

~~$u = 2-t$~~

~~$\frac{du}{dt} =$~~

$u = 2-t$ $\qquad y = \ln\left(\frac{2}{u}\right)$

$\frac{du}{dt} = -1$ $\qquad \frac{dy}{du} = \frac{-1}{u}$

(for y: Lt $z = \frac{2}{u}$ $\qquad y = \ln(z)$

$\frac{dz}{du} = -\frac{2}{u^2}$ $\qquad \frac{dy}{dz} = \frac{1}{z}$

$\frac{dy}{du} = -\frac{2}{u^2} \times \frac{1}{\frac{2}{u}} = -\frac{2}{u^2} \times \frac{u}{2} = \frac{-1}{u}$)

$\frac{dy}{dt} = \frac{-1}{u} \times -1 = \frac{1}{u} = \frac{1}{2-t}$

$$G'_X(t) = \frac{1}{\ln 2} \times \frac{1}{2-t}$$

~~G~~ for $G''_X(t)$

~~Let~~ $u = 2-t$ $\qquad y = \frac{1}{u}$

$\frac{du}{dt} = -1 \qquad \frac{dy}{du} = -\frac{1}{u^2}$

$$\frac{dy}{dt} = -1 \times -\frac{1}{u^2} = \frac{1}{(2-t)^2}$$

$$\therefore G''_X(t) = \frac{1}{\ln 2}\left(\frac{1}{(2-t)^2}\right)$$

$$Var(X) = G''_X(1) + G'_X(1) - \left[\cancel{\{}G'_X(1)^2\right]$$

$$= \frac{1}{\ln 2}\left(\frac{1}{(1)^2}\right) + \frac{1}{\ln 2}\left(\frac{1}{1}\right) - \left[\frac{1}{\ln 2^2}\left(\frac{1}{1}\right)^2\right)$$

$$= \frac{2}{\ln 2} - \frac{1}{(\ln 2)^2} = \frac{2}{\ln 2} - \frac{1}{2\ln 2}$$

$$= \frac{3}{2\ln 2}$$

7/12

Examiner Comments

Part (a) is correct (B1).

In part (b) the differentiation is, eventually carried out correctly and the correct formula for variance is used (M1A1A1M1A1M1). This student performed a common error though when dealing with $\left(\frac{1}{\ln 2}\right)^2$ and thought this was $\frac{1}{2\ln 2}$ which led to an incorrect final answer (A0).

In part (c) there is a mention of the need for the coefficient of t^3 but no attempt to find a suitable series expansion (M0A0M0A0).

Student Response C

a) $G_x(1)=1$

$k\ln\left(\frac{2}{2-1}\right)=1$

$k\ln\left(\frac{2}{1}\right)=1$

$k\ln 2=1$

$k=\frac{1}{\ln 2}$

b) $G_x(t)=-k\ln(1-\frac{1}{2}t)$

$G'_x(t)=\frac{-k(-\frac{1}{2})}{1-\frac{1}{2}t}=\frac{k}{2-t}=k(2-t)^{-1}$

$G''_x(t)=-k(2-t)^{-2}(-1)$

$=\frac{k}{(2-t)^2}$

$\left[G'_x(1)=\frac{\left(\frac{1}{\ln 2}\right)}{2-1}=\frac{\frac{1}{\ln 2}}{1}=\frac{1}{\ln 2}\right]$

$\left[G''_x(1)=\frac{\left(\frac{1}{\ln 2}\right)}{(2-t)^2}=\frac{\left(\frac{1}{\ln 2}\right)}{1^2}=\frac{1}{\ln 2}\right]$

$Var(X)=G''_x(1)+G'_x(1)-(G'_x(1))^2$

$=\frac{1}{\ln 2}+\frac{1}{\ln 2}-\left(\frac{1}{\ln 2}\right)^2$

$=\frac{2}{\ln 2}-\frac{1}{\ln 2\ln 2}$

$=\frac{2\ln 2-1}{(\ln 2)^2}$

c) ~~$\ln(2-t)$~~

~~$\ln(1-\frac{1}{2}t)+x=\ln(1+t)$~~

~~$x=\ln\left(\frac{1+t}{1-\frac{1}{2}t}\right)$~~

$G_x(0)=0$

$G_x(t)=G(0)+tG'(0)+\frac{t^2}{2!}G''(0)\boxed{+\frac{t^3}{3!}G'''(0)}$

$P(X=3)=\frac{G'''_x(0)}{3!}$

$G''_x(t)=k(2-t)^{-2}$

$G'''_x(t)=2k(2-t)^{-3}=\frac{2k}{(2-t)^3}$ $\quad G'''_x(0)=\frac{\left(\frac{2}{\ln 2}\right)}{2^3}=\frac{2}{8\ln 2}$

$P(X=3)=\frac{2}{8\ln 2\times 3!}=\underline{0.0601}$

12/12

Examiner Comments

Part (a) is correct (B1).

Part (b) is correct (M1A1A1M1A1M1A1).

In part (c) there is a correct 3rd derivative (M1A1) and a correct statement using this to find the required probability (M1) and this leads to the correct final answer on the next page (A1). The approach using Maclaurin's expansion was by far the most common and successful one used.

Exemplar Question 7

back to Contents Page

7. A spinner can land on red or blue. When the spinner is spun, there is a probability of that it lands on blue. The spinner is spun repeatedly.

The random variable *B* represents the number of the spin when the spinner first lands on blue.

(*a*) Find (i) $\mathrm{P}(B = 4)$

(ii) $\mathrm{P}(B \leq 5)$

(4)

(*b*) Find $\mathrm{E}(B^2)$

(3)

Steve invites Tamara to play a game with this spinner.

Tamara must choose a colour, either red or blue.

Steve will spin the spinner repeatedly until the spinner first lands on the colour Tamara has chosen. The random variable *X* represents the number of the spin when this occurs.

If Tamara chooses red, her score is e^X

If Tamara chooses blue, her score is X^2

(*c*) State, giving your reasons and showing any calculations you have made, which colour you would recommend that Tamara chooses.

(5)

(Total for Question 7 is 12 marks)

Mean Score 6.43 out of 12

Examiner Comments

Part (a) was answered very well with most using the correct geometric distribution but some used $1 – \mathrm{P}(B \geq 4)$ rather than $1 – \mathrm{P}(B > 5)$.

Part (b) was usually answered very well too with the most common error being to use $\mathrm{E}(B^2) = [\mathrm{E}(B)]^2$.

Part (c) was the most challenging question on the paper. Many obtained the first mark for identifying the correct distribution for the number of the spin when it first lands on red. The main problem was that students didn't know how to find $\mathrm{E}(\mathrm{e}^X)$ with $\mathrm{e}^{\mathrm{E}(X)}$ being the most common incorrect approach. Some tried to argue from comparing the graphs of $y = \mathrm{e}^x$ and $y = x^2$ but there was no consideration of appropriate probabilities and some numerical approaches floundered in the same way. A small number found the correct probability generating function $\mathrm{G}(t)$ and then gave the very succinct solution $\mathrm{G}(\mathrm{e})$ but this was rare! Over 5% though did find the correct value of $\mathrm{E}(\mathrm{e}^X)$ and after comparing this with $\mathrm{E}(B^2)$ were able to secure all the marks for this question.

Mark Scheme

Qu	Scheme	Marks	AO
7(a)(i)	[B~Geo($\frac{1}{3}$)] P(B = 4) = $\left(\frac{2}{3}\right)^3 \times \frac{1}{3}$	M1	3.3
	$= \underline{\frac{8}{81}}$	A1	1.1b
(ii)	P($B \leqslant 5$) = 1 – P(B > 5) <u>or</u> $1-\left(\frac{2}{3}\right)^5$	M1	2.1
	$= \underline{\frac{211}{243}}$	A1	1.1b
		(4)	
(b)	$\text{E}(B^2) = \text{Var}(B) + [\text{E}(B)]^2$	M1	2.1
	From formula booklet: E(B) = $\frac{1}{\frac{1}{3}}$ = 3 and Var(B) = $\frac{1-\frac{1}{3}}{\left(\frac{1}{3}\right)^2}$ = 6	B1	1.1b
	So E(B^2) = 6 + 9 = $\underline{\mathbf{15}}$	A1	1.1b
		(3)	
(c)	[Let R = no. of the spin when it first lands on red] X = R~Geo($\frac{2}{3}$)	M1	3.3
	Require E(e^X) = $\sum_{x=1}^{\infty} \text{e}^x \left(\frac{1}{3}\right)^{x-1} \frac{2}{3}$	M1	3.1a
	$= \frac{2\text{e}}{3} \sum_{x=1}^{\infty} \left(\frac{\text{e}}{3}\right)^{x-1}$	M1	2.1
	$= \frac{2\text{e}}{3} \times \frac{1}{1-\frac{\text{e}}{3}}$ <u>or</u> $\frac{2\text{e}}{3-\text{e}}$	A1	1.1b
	E(e^X) = 19.297… {> 15 = E(B^2)} so Tamara should **choose red** since it has the greater expected score	A1	2.2a
		(5)	
		(12 marks)	

	Notes
(a)(i)	M1 for selecting the correct model i.e. Geo(p) (May be implied by a correct expression) A1 for $\frac{8}{81}$ (= 0.098765… accept awrt 0.0988)
(ii)	M1 for a suitable strategy to use the geometric model to find a correct expression A1 for $\frac{211}{243}$ (= 0.868312…accept awrt 0.868)
(b)	M1 for a suitable strategy to find E(B^2) [allow $\text{G}''(1)+\text{G}'(1)$] B1 for use of the correct formulae to find E(B) = 3 <u>and</u> Var(B) = 6 <u>or</u> $\text{G}''(1) = 12$ A1 for 15
SC	**Formula for E(B^2)** Allow M1B1A0 for $\text{E}(B^2) = \frac{2-p}{p^2}$ (o.e.)
(c)	1st M1 for choosing a suitable geometric model (sight of Geo($\frac{2}{3}$) or at least 3 correct probabilities) 2nd M1 for realising the need for appropriate expected value and using E(g(X)) [Need sum and f(x)] NB simply finding $\text{e}^{\text{E}(X)} = \text{e}^{1.5}$ = awrt 4.48 is M0 and probably no more marks. 3rd M1 for a suitable strategy to turn the expression into a sum that can be found 1st A1 for correct use of sum to infinity of geometric series 2nd A1 for interpreting the outcome of the calculations in terms of a solution to the problem must choose red and see the awrt 19.3 (and allow ft of their E(B^2) < 19)

Examiner Comments

In part (c) the first mark for stating the correct model for the number of the spin when it first lands on red was readily accessible but the “doorkeeper” to the remaining marks was the next M mark which required a complete expression (a sum and a function of x or a clear attempt to find the infinite sum) for $E(e^X)$.

Student Response A

a(i) $B \sim Geo(1/3)$

$p(B = 1/3)$

$p(B=3) = 1/3 \times 2/3^2 = \frac{4}{27} = \underline{0.148}$

ii) $p(B \le 5) = 1 - q^5 = 1 - 2/3^5 = 0.868$

b) $E(B = \frac{1}{1/3} = 3$

$E(B^2) = E(B) \times E(B) = 3 \times 3 = 9$

c) ~~$e^x = 1 + x + \frac{x^3}{2} + \ldots$~~

$P(\text{Red}) = 2/3$

$P(\text{Blue}) = 1/3$

$E(R) =$ [illegible] $E(R) = 1/2/3 = 1.5$

$E(B) = 1/1/3 = 3$

$E(B) =$ [illegible]

c) so expected score for Red $= e^{1.5}$

Expected score for blue $= 3^2 = 9$

Blue

I would recommend that the chosen blue as the expected score is higher $(9 > e^{1.5})$

3/12

Examiner Comments

In part (a) the correct Geometric model is stated (M1) but the probability in (i) is incorrect (A0) but the formula and answer are correct in part (ii) (M1A1).

In part (b) they give a correct value for E(*B*) but the B mark requires a correct variance as well which they never give (B0). They think that $E(B^2) = E(B) \times E(B)$ (M0) and give the common incorrect answer of 9 (A0).

In part (c) they do not explicitly state that $R \sim \text{Geo}(\frac{2}{3})$ is the suitable model to be used (M0) and simply think that $E(e^R) = E(e^{E(R)})$ and no further marks are possible.

Student Response B

a) ~~$X \sim NB(4, 1/3)$~~

$B \sim \text{Geo}(1/3)$

i) $P(B = 4) = \frac{1}{3}\left(\frac{2}{3}\right)^3$

$= 8/81 \approx 0.0988$

ii) $P(B \leq 5) = 1 - q^x$

$= 1 - (2/3)^5 = \frac{211}{243} \approx 0.8683$

b) $E(B) = 1/p$

$\text{Var}(B) = \frac{1-p}{p^2} = E(B^2) - (E(B))^2$

$\text{Var}(B) + (E(B))^2 = E(B^2)$

$\therefore E(B^2) = \frac{1-p}{p^2} + \frac{1}{p^2} = \frac{2-p}{p^2}$

c) ~~Tamara's expected~~

The number of spins ~~it takes~~ Tamara can expect it to take to land on blue is $1/\frac{1}{3} = 3$

$\therefore$ she can expect an average score of $3^2 = 9$ if she picks blue

The number of spins tamara can expect it to take to land on red is $1/(2/3) = 3/2$, so she can expect an average score of $e^{3/2}$ if she picks red

Since $e^{3/2} < 9$, I would recommend that Tamara picks blue.

6/12

Examiner Comments

In part (a) both parts are fully correct (M1A1M1A1).

In part (b) there is a suitable strategy given (M1) and a correct formula found for E(B2) (B1) but this is not evaluated (A0). A special case was detailed on the mark scheme to enable 2 of the 3 marks to be awarded in this case.

In part (c) a correct model for the number of the spin when the first red occurs (R) is not explicitly given (M0) and they simply compare eE(R) with E(B)2 which is a common error and does not score any further marks (M0M0A0A0).

Student Response C

7.a)(i) $B \sim \text{Geo}(\frac{1}{3})$

$P(B=4) = (\frac{2}{3})^3(\frac{1}{3})$

$= \frac{8}{81} = 0.09877$

$= 0.0988$

(ii) $P(B \leq 5) = 1 - P(X>5)$

$= 1 - (\frac{2}{3})^5$

$= \frac{211}{243} = 0.8683$

$= 0.868$

b) $\text{Var}(B) = E(B^2) - E(B)^2$

$E(B^2) = \text{Var}(B) + E(B)^2$

$E(B) = \frac{1}{\frac{1}{3}} = 3$

$\text{Var}(B) = \frac{2}{3} \div (\frac{1}{3})^2$

$= \frac{2}{3} \times 9$

$= 6$

$E(B^2) = 6 + 3^2 = 6 + 9 = 15$

c) Red $X \sim \text{Geo}(\frac{2}{3})$

$E(X) = 1.5$

$E(e^x) =$

e^x	e^1	e^2	e^3	e^4	e^5	....
$P(X=x)$	$(\frac{2}{3})$	$(\frac{2}{3})(\frac{1}{3})$	$\frac{2}{3}(\frac{1}{3})^2$	...	...	....

$E(e^x) = e(\frac{2}{3}) + e^2(\frac{2}{3})(\frac{1}{3}) + e^3(\frac{2}{3})(\frac{1}{3})^2 + \ldots.$

$= e\frac{2}{3}(1 + e(\frac{1}{3}) + e^2(\frac{1}{3})^2 + \ldots)$

$= e(\frac{2}{3})\left(\frac{1}{1-\frac{e}{3}}\right) = \frac{2e}{3} \times \frac{1}{\frac{3-e}{3}} = \frac{2e}{\cancel{3}} \times \frac{\cancel{3}}{3-e} = \frac{2e}{3-e}$

$= 19.298$

$= 19.3$

Blue $X \sim \text{Geo}(\frac{1}{3})$

$E(X) = 3$

$E(X^2) = 15$

As the expected value of the score if Tamara chooses red is greater than ~~that is~~ the expected score of choosing blue, $E(e^x) > E(X^2)$

$19.298 > 15$

I would suggest Tamara to choose red.

12/12

Examiner Comments

In part (a) this is fully correct (M1A1M1A1).

In part (b) a correct formula (strategy) is given (M1) and the correct values for the mean and variance are stated (B1) leading to the correct answer (A1).

In part (c) the correct Geometric model is stated (M1) and a correct expression for $\mathrm{E}(\mathrm{e}^X)$ is implied by the correct infinite sum indicated (M1). After taking out a suitable factor they identify that this sum is a geometric series and use the sum to infinity formula (M1) which gives the correct answer of $\frac{2\mathrm{e}}{3-\mathrm{e}}$ (A1). This is evaluated and compared with the answer to part (b) and a correct conclusion (that Tamara should choose red) is made (A1).

A Level Further Mathematics – Further Statistics 2 (9FM0 4B)

Exemplar Question 1

back to Contents Page

1 A machine is set to fill pots with yoghurt such that the mean weight of yoghurt in a pot is 505 grams.

To check that the machine is working properly, a random sample of 8 pots is selected. The weight of yoghurt, in grams, in each pot is as follows

508 510 500 500 498 503 508 505

Given that the weights of the yoghurt delivered by the machine follow a normal distribution with standard deviation 5.4 grams,

(*a*) find a 95% confidence interval for the mean weight, μ grams, of yoghurt in a pot. Give your answers to 2 decimal places.

(4)

(*b*) Comment on whether or not the machine is working properly, giving a reason for your answer.

(1)

(*c*) State the probability that a 95% confidence interval for μ will not contain μ grams.

(1)

(*d*) Without carrying out any further calculations, explain the changes, if any, that would need to be made in calculating the confidence interval in part (*a*) if the standard deviation was unknown. Give a reason for your answer.
You may assume that the weights of the yoghurt delivered by the machine still follow a normal distribution.

(2)

(Total for Question 1 is 8 marks)

Mean Score 5.3 out of 8

Examiner Comments

This question is assessing the student’s understanding of the concept of a confidence interval and its interpretation as well as calculating a confidence interval for a normal mean with variance known and with variance unknown.

Part (a) requires the students to calculate a confidence interval for a normal mean, with variance known.

Parts (b) and (c) assess the students understanding of a confidence interval.

Part (d) requires the students to understand the differences in calculating a confidence interval for a normal mean with variance known and a confidence interval for a normal mean with variance unknown. It also requires an understanding of why certain methods should be used.

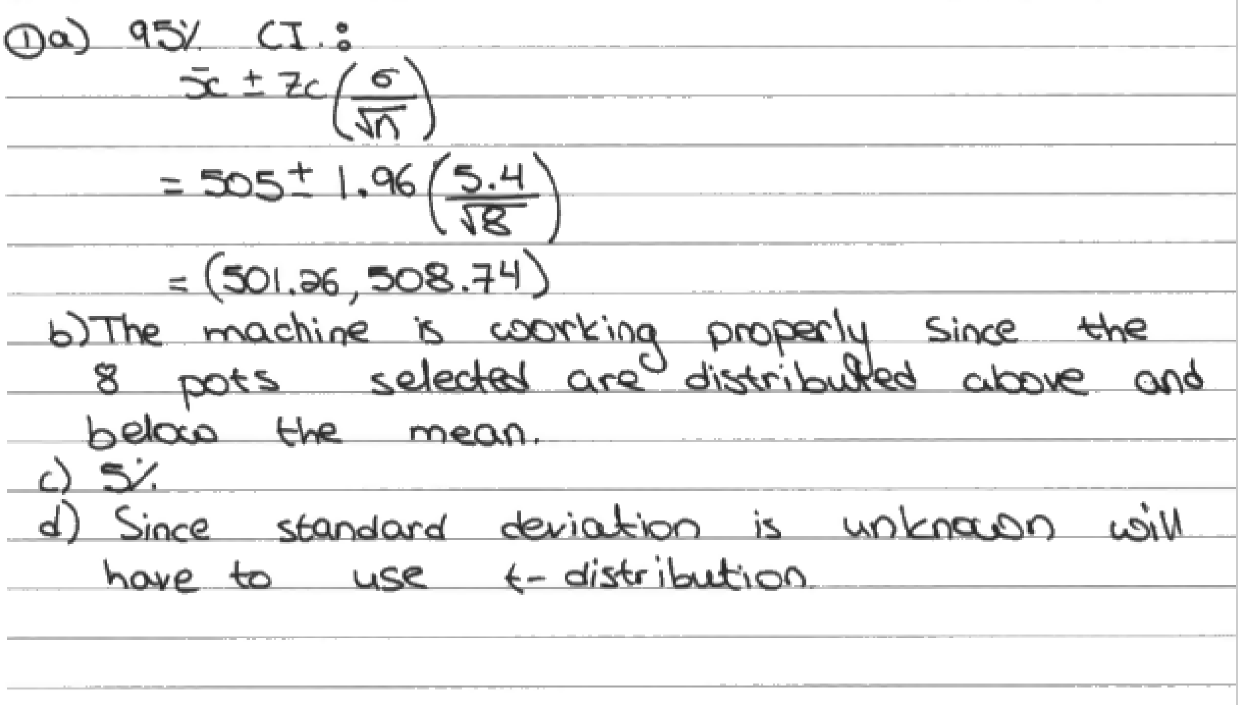
Mark Scheme

Question	Scheme	Marks	AOs
1(a)	Mean = 504	B1	1.1b
	1.96	B1	3.3
	$504 \pm \frac{5.4}{\sqrt{8}} \times "1.96"$	M1	2.1
	(500.258, 507.742)	A1	1.1b
		(4)	
(b)	**505** is in the confidence **interval** therefore there is evidence that the machine is **working** properly	B1ft	2.2b
		(1)	
(c)	5% oe	B1	1.1b
		(1)	
(d)	s needs to be used instead of σ **and** a t-value instead of the z value	B1	3.3
	since the sample is small therefore you can't use the normal distribution	B1	3.5b
		(2)	
			(8 marks)

Notes:

(a)	**B1**	504 may be seen in part (b)
	B1	For realising a normal distribution must be used as a model and finding the correct value 1.96
	M1	For $504 \pm \frac{5.4}{\sqrt{8}} \times "z \text{ value}"$. $\lvert z \rvert > 1$ May be implied by a correct CI
	A1	awrt 500.26 and 507.74 NB using t gives 500.29 and 507.71
(b)	**B1ft**	Drawing a correct inference (ft) using their answer to part (a) and the 505 from the question. Reason must be given. Ignore incorrect non – contextual
(c)	**B1**	5%
(d)	**B1**	create new model by using s and t. Allow if state use CI $\mu \pm \frac{s}{\sqrt{n}} \times "t"$ or use s = 4.44 and t = 2.365
	B1	For recognising that the sample is small

Student Response A



①a) 95% C.I.:

$$\bar{x} \pm z_c\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$= 505 \pm 1.96\left(\frac{5.4}{\sqrt{8}}\right)$$

$$= (501.26, 508.74)$$

b) The machine is working properly since the 8 pots selected are distributed above and below the mean.

c) 5%

d) Since standard deviation is unknown will have to use t- distribution.

2/8

Examiner Comments

The mean 504 cannot be seen in part (a) or part (b). The correct critical value is used. However, the confidence interval is calculated using 505 rather than the mean of the sample. **B0B1M0A0**

In part (b) 505 is not seen. **B0ft**

In part (c) 5% is correct. **B1**

In part (d) the explanation includes the fact that *t* needs to be used but there is no indication that *s* needs to be used as well. **B0**

No reason for why these need to be used is given. **B0**

Student Response B

a) $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$

$\bar{x} = 504 \quad z = 1.96 \quad \sigma = 5.4 \quad \sqrt{n} = \sqrt{8}$

$504 \pm 1.96 \times \frac{5.4}{\sqrt{8}}$

$(500, 508)$

b) the machine is working properly as the mean weight of 505 grams lies within the confidence interval

c) 5%

d) there would be changes as we would need to use the values in the t_{n-1} distribution instead of using the normal distribution

5/8

Examiner Comments

In part (a) the correct calculation for the confidence interval is given but the answer is not given to the required accuracy. **B1B1M1A0**

In part (b) 505 is used and a correct conclusion is given with the required words seen. **B1ft**

In part (c) 5% is correct. **B1**

In part (d) the explanation includes the fact that *t* needs to be used but there is no indication that *s* needs to be used as well. **B0**

No reason for why these need to be used is given. **B0**

Student Response C

a) $n = 8$ by calculator

$\Sigma x = 4032$

$\Sigma x^2 = 2032,266$

$\bar{x} = \frac{\Sigma x}{n} = 504$

$s^2 = \frac{\Sigma x^2 - n\bar{x}^2}{n-1} = 17.71$

$[500.26, 507.74]$

95% C.I for μ

$z = 1.96$

$\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$

$= 504 \pm 3.742$ = ~~[500.23, 507.7] approx~~

~~[500.3, 507.7]~~

b) 505 lies in the interval so at a significance of 5% there is ~~ev~~ sufficient evidence to conclude the machine is working properly

c) ~~95~~ 0.05

d) Instead of using the critical value ~~z~~ from the standard normal distribution you would use $t_{(n-1)}(0.025)$ which in this case would be 2.365

and ~~instead~~ instead of using σ you would use the unbiased estimate of standard deviation $\sqrt{s^2}$

formula for confidence limit would become

$\bar{x} \pm 2.365 \times \frac{s}{\sqrt{n}}$

7/8

Examiner Comments

In part (a) the correct confidence interval is given. **B1B1M1A1**

In part (b) 505 is used and a correct conclusion is given with the required words seen. **B1ft**

In part (c) 0.05 is equivalent to 5%. **B1**

In part (d) the explanation includes the fact that t needs to be used as well as s. **B1**
No reason for why these need to be used is given. **B0**

Exemplar Question 2

back to Contents Page

2 A large field of wheat is split into 8 plots of equal area. Each plot is treated with a different amount of fertiliser, grams /m^2. The yield of wheat, *w* tonnes, from each plot is recorded. The results are summarised below.

$$\sum f = 28 \qquad \sum w = 303 \qquad \sum w^2 = 13\,447 \qquad S_{ff} = 42 \qquad S_{fw} = 269.5$$

(*a*) Calculate the product moment correlation coefficient between *f* and *w*

(2)

(*b*) Interpret the value of your product moment correlation coefficient.

(1)

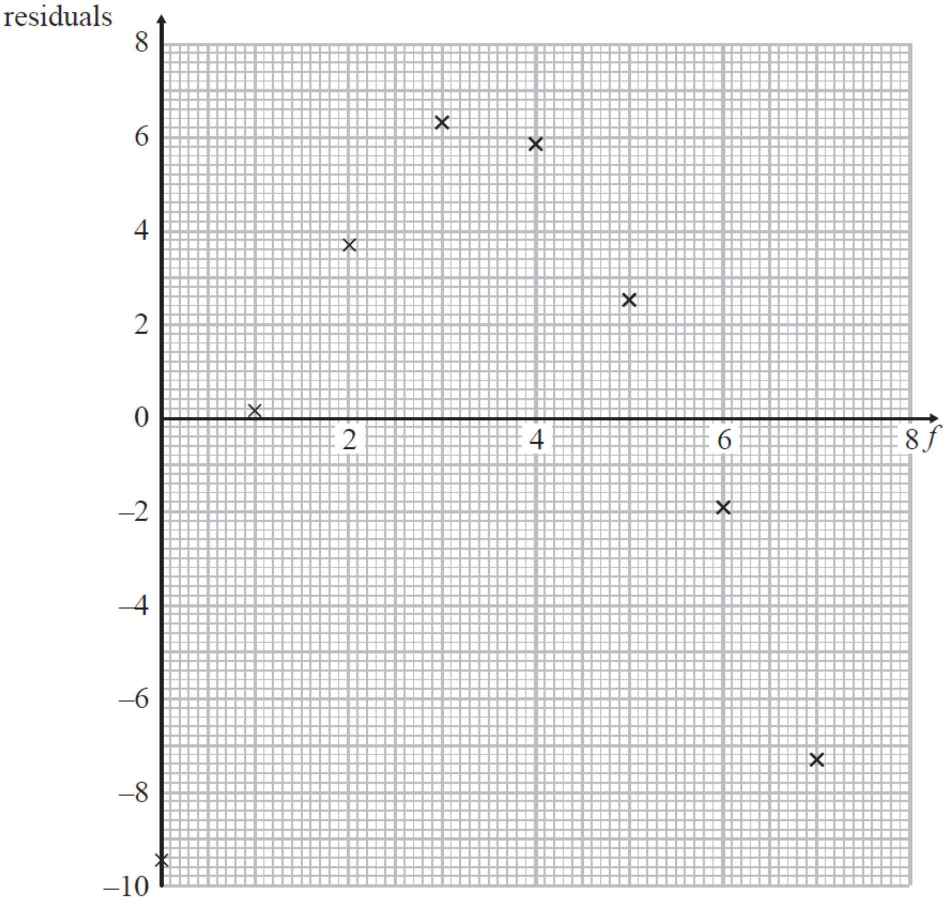
(*c*) Find the equation of the regression line of *w*on *f* in the form $w = a + bf$

(3)

(*d*) Using your equation, estimate the decrease in yield when the amount of fertiliser decreases by 0.5 grams /m2

(1)

The residuals of the data recorded are calculated and plotted on the graph below.



(*e*) With reference to this graph, comment on the suitability of the model you found in part (*c*).

(2)

(*f*) Suggest how you might be able to refine your model.

(1)

(Total for Question 2 is 10 marks)

Mean Score 7.2 out of 10

Examiner Comments

This question is assessing the student's understanding of the product moment correlation coefficient (3.1), regression lines (1.1) and residuals (1.2).

Part (a) requires the students to calculate the product moment correlation coefficient.

In part (b) the students are asked to interpret the product moment correlation coefficient. The demand "Interpret" indicates that the answer needs to be in the context of the question.

Part (c) requires the students to find the equation of a regression line.

In part (d) the student is required to use their regression line to estimate a value.

Part (e) requires the students to refer to the graph to comment on the suitability of the model. The main misconception was that the residuals needed to be close to zero for the model to be suitable rather than they need to be randomly scattered. A minority of students stated that the residuals did not add up to 1 possibly having not noticed the residual for $f = 0$.

In part (f) very few students realised that the graph indicated that a nonlinear model would be more appropriate.

Mark Scheme

Question	Scheme	Marks	AOs
2(a)	$S_{ww}=13447-\frac{303^2}{8}=1970.875$		
	$r=\frac{269.5}{\sqrt{42\times1970.875}}$	M1	1.1b
	$r=0.9367...$awrt 0.937	A1	1.1b
		(2)	
(b)	As the amount of **fertiliser** increases the **yield** increases	B1	3.2a
		(1)	
(c)	$b=\frac{269.5}{42}\left[=6.41666...\right]$	M1	3.3
	$a=\frac{303}{8}-'b'\frac{28}{8}\left[=15.41666...\right]$	M1	1.1b
	$w=15.4+6.42f$	A1	1.1b
		(3)	
(d)	3.21 tonnes	B1ft	1.1b
		(1)	
(e)	The **residual** plot is close to an ’n’ shape or the <u>**residuals**</u> appear not to be **randomly** scattered	M1	2.4
	The model in part (c) is unlikely to be suitable	A1	2.2b
		(2)	
(f)	Fit a curve rather than a line	B1	3.5c
		(1)	
		(10 marks)	

Notes:

(a)	**M1** **A1**	Complete correct method for finding r for awrt 0.937
(b)	**B1**	Correct contextual statement
(c)	**M1**	For use of a correct model ie a correct expression for b
	M1	For use of a correct model ie a correct expression (ft) for a
	A1	For a correct model $w=15.4+6.42f$ with awrt 15.4 and awrt 6.42
(d)	**B1ft**	awrt 3.21 condone -3.21
(e)	**M1**	Explaining a reason for their conclusion e.g. there is **a pattern/trend** in the **residuals** Do not accept residuals not close to zero
	A1	concluding it is not valid oe
(f)	**B1**	A comment about not using a linear line e.g. use a quadratic model, logarithmic graph exponential

Student Response A

(a) $r = \dfrac{S_{fw}}{\sqrt{S_{ff}S_{ww}}}$

$S_{ff} = 42$

$S_{fw} = 269.5$

$S_{ww} = \sum w^2 - \dfrac{(\sum w)^2}{8}$

$= 13447 - \dfrac{303^2}{8} = 1970.875$.

$\Rightarrow r = \dfrac{269.5}{\sqrt{42 \times 1970.875}} = \underline{0.937\ (3s.f.)}$.

(b) ~~[illegible]~~ $r = 0.937$ is a positive correlation.

~~The amount of [illegible]~~ .

When the amount of fertiliser ~~[illegible]~~ increases, the yield of wheat also increases.

(c) When $y = 2x + b$,

~~[illegible]~~ The regression coefficient of y on x .

~~[illegible]~~ $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{S_{fw}}{S_{ff}} = \dfrac{269.5}{42}$ ~~[illegible]~~

$= 6.42\ (3s.f.)$

$a =$

~~[illegible]~~ $\Rightarrow w = \quad + 6.42f$

~~[illegible]~~ .

(d) $w = a + 6.42f$

When $f = 0.5$, $w = a + 6.42 \times 0.5$

$w = a + 3.21$

(e) shows the gradient

4/10

Examiner Comments

In part (a) an awrt 0.937 is given. **M1A1**

In part (b) a correct answer in context has been given. **B1**

In part (c) the correct expression for*b* has been given. No expression for *a* has been given and the equation of the regression line is incorrect. **M1M0A0**

In part (d) the student has attempted to find the value of *w* when *f* = 0.5 rather than the decrease in yield using their value of *b*. **B0ft**
the correct value for the decrease in yield is given. **B0ft**

In part (e) it does not say why the graph suggests a non-linear model. **M0A0**

In part (f) No answer is given. **B0**

Student Response B

w

$$S_{ww} = \Sigma w^2 - \frac{(\Sigma w)^2}{n}$$

$$= 13447 - \frac{(303)^2}{8}$$

$$= 1970,875$$

$$r = \frac{S_{fw}}{\sqrt{S_{ww} S_{ff}}} = \frac{269.5}{\sqrt{1970.875 \times 42}} = 0.937$$

b). There's a strong positive ~~linear~~ correlation between f and ~~to~~ $\overset{w}{}$.

c). $b = \frac{S_{wf}}{S_{ff}}$

$$= \frac{269.5}{42}$$

$$= \frac{77}{12}. = ~~0~~ 6.42.$$

$\bar{w} = a + b\bar{f}$

$\frac{303}{8} = a + \frac{77}{12} \times \frac{28}{8}$

$a = \frac{185}{12}$

$= 15.4$

$\therefore w = \frac{185}{12} + \frac{77}{12} f$

$w = 15.4 + 6.42f$

d) when $f = -0.5$

$$w = \frac{185}{12} - \frac{77}{12} \times 0.5$$

$$= \frac{293}{24}.$$

$\frac{185}{12} - \frac{293}{24} = 3.21$ (tonnes). $\therefore$ It decreases by 3.21 tonnes.

e) The graph does not suggest a linear correlation

$\therefore$ The regrassion line is not suitable.

f) Use spearman's rank model since the correlation is not ~~linear~~.

6/10

Examiner Comments

In part (a) an awrt 0.937 is given. **M1A1**

In part (b) the answer is not given in context. **B0**

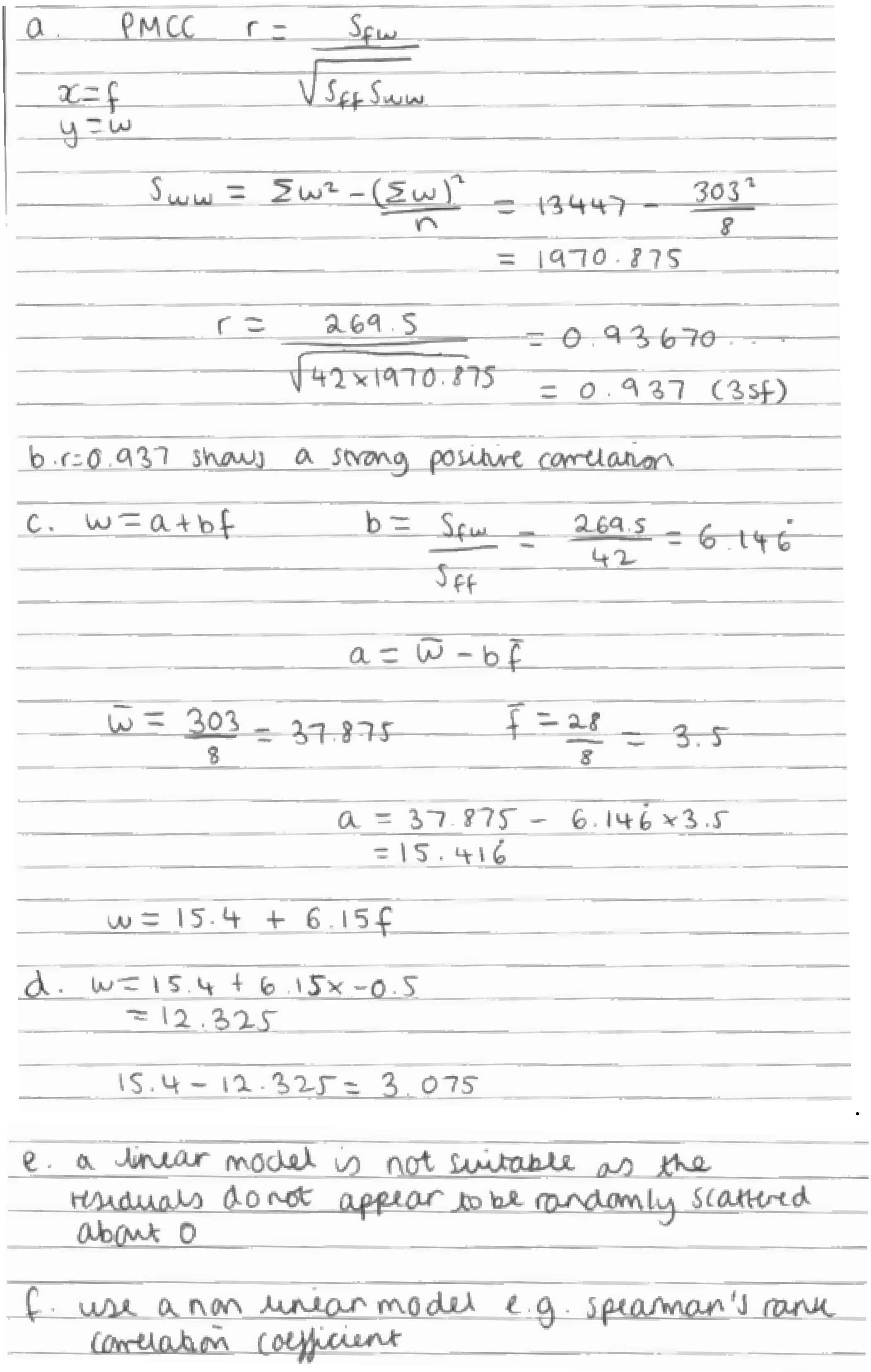
In part (c) a correct equation of the regression line is given. **M1M1A1**

In part (d) the correct value for the decrease in yield is given. **B1ft**

In part (e) it does not say **why** the graph suggests a non-linear model. **M0A0**

In part (f) A comment about not using a linear line has not been given. **B0**

Student Response C



a. PMCC $r = \frac{S_{fw}}{\sqrt{S_{ff}S_{ww}}}$

$x=f$
$y=w$

$S_{ww} = \Sigma w^2 - \frac{(\Sigma w)^2}{n} = 13447 - \frac{303^2}{8}$
$= 1970.875$

$r = \frac{269.5}{\sqrt{42\times1970.875}} = 0.93670\ldots$
$= 0.937$ (3sf)

b. $r=0.937$ shows a strong positive correlation

c. $w=a+bf$ $b = \frac{S_{fw}}{S_{ff}} = \frac{269.5}{42} = 6.14\dot{6}$

$a = \bar{w} - b\bar{f}$

$\bar{w} = \frac{303}{8} = 37.875$ $\bar{f} = \frac{28}{8} = 3.5$

$a = 37.875 - 6.14\dot{6}\times3.5$
$= 15.41\dot{6}$

$w = 15.4 + 6.15f$

d. $w = 15.4 + 6.15\times -0.5$
$= 12.325$

$15.4 - 12.325 = 3.075$

e. a linear model is not suitable as the residuals do not appear to be randomly scattered about 0

f. use a non linear model e.g. spearman's rank correlation coefficient

8/10

Examiner Comments

In part (a) an awrt 0.937 is given. **M1A1**

In part (b) the answer is not given in context. **B0**

In part (c) a correct expression for *b* and *a* is given but the value of *b* is incorrect. **M1M1A0**

In part (d) the student has found the decrease in yield using their value of *b*. **B1ft**

In part (e) a correct answer with a suitable reason has been given. **M1A1**

In part (f) using a nonlinear model is equivalent to fit a curve rather than a line reason has been given. **B1**

Exemplar Question 3

back to Contents Page

3 Yin grows two varieties of potato, plant *A* and plant *B*. A random sample of each variety of potato is taken and the yield, x kg, produced by each plant is measured. The following statistics are obtained from the data.

	Number of plants	$\sum x$	$\sum x^2$
A	25	194.7	1637.37
B	26	227.5	2031.19

(*a*) Stating your hypotheses clearly, test, at the 10% significance level, whether or not the variances of the yields of the two varieties of potato are the same.

(7)

(*b*) State an assumption you have made in order to carry out the test in part (*a*).

(1)

(Total for Question 3 is 8 marks)

Mean Score 6.6 out of 8

Examiner Comments

This question is assessing the student's ability to test whether or not two independent random samples are from Normal populations with equal variances.

Part (a) requires the students to use the data given to complete the hypothesis test.

In part (b) the students need to demonstrate they understand what assumptions have been made in order to carry out the test in part (a). Many students knew that something had to be normally distributed, but few realised that it was the yield that was assumed to be normally distributed.

Mark Scheme

Question	Scheme	Marks	AOs
3(a)	$\mathrm{H}_0 : \sigma_A^2 = \sigma_B^2,\ \mathrm{H}_1 : \sigma_A^2 \neq \sigma_B^2$	B1	2.5
	$s_A^2 = \frac{1}{24}\left(1637.37 - 25 \times \left(\frac{194.7}{25}\right)^2\right) = 5.0436$	M1 A1	2.1 1.1b
	$s_B^2 = \frac{1}{25}\left(2031.19 - 26 \times \left(\frac{227.5}{26}\right)^2\right) = 1.6226$	A1	1.1b
	$\frac{s_A^2}{s_B^2} = 3.108...$	M1	3.4
	critical values upper tail $\mathrm{F}_{24,25} = 1.96$	B1	1.1b
	There is evidence that the two **variances** are different.	A1ft	2.2b
		(7)	
(b)	The **yields** are normally distributed.	B1	1.2
		(1)	
		(8 marks)	

Notes:

(a)	**B1**	both hypotheses correct using σ **or** σ^2
	M1	Using a correct method for either s_A^2 or s_B^2. May be implied by a correct value
	A1	awrt 5.04
	A1	awrt 1.62
	M1	Using the F-distribution as the model eg $\frac{s_A^2}{s_B^2}$ $\left(\text{allow } \frac{s_B^2}{s_A^2} [= 0.321...]\right)$
	B1	awrt1.96 or 0.506 must match their method
	A1ft	Drawing a correct inference following through their CV and value for $\frac{s_B^2}{s_A^2}$ or $\frac{s_A^2}{s_B^2}$ Allow $\sigma_B^2 \neq \sigma_A^2$ Allow standard deviation instead of Var .Do not allow $\sigma_B^2 = \sigma_A^2$
(b)	**B1**	recalling the fact that the variable **yield** needs to be normally distributed

S0tudent Response A

(a) $H_0 : \sigma_A^2 = \sigma_B^2$ s.l. ⇒ 0.005

$H_1 : \sigma_A^2 \neq \sigma_B^2$

critical value :

Var A =

(b) Two plant A and plant B are grown in the same condition.

1/8

Examiner Comments

In part (a) the hypotheses are correct, but no further progress has been made. **B1M0A0A0M0B0A0**

In part (b) the assumption needed to use the test has not been given. **B0**

Student Response B

a) $H_0: \sigma_a = \sigma_b$

$H_1: \sigma_a \neq \sigma_b$ 2-tailed test ∴ 5% critical value.

$S_a^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$

$\frac{S_a^2}{S_b^2} \sim F_{n_a-1, n_b-1}\overset{0.05}{(\text{[illegible]})}$

$\rightarrow S_a^2 = 4.841856$

$F_{24,25}\overset{0.05}{(\text{[illegible]})} = \boxed{2.62}$

$S_b^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$

$\rightarrow S_b^2 = 1.5601$

$\frac{S_a^2/\sigma_a^2}{S_b^2/\sigma_b^2} = \frac{S_a^2}{S_b^2} = \boxed{3.103371281}$

3.1033 > 2.62 ∴ sufficient evidence to reject Ho. the Variances are not the same at the 10% level.

b) assumed that the data came from a normally [illegible] distributed population.

3/8

Examiner Comments

In part (a) the hypotheses are correct. An incorrect method has been used to find s_A^2 and s_B^2. The F-distribution model has been used with their s_A^2 and s_B^2 but the critical value is incorrect. As the previous method mark has been awarded this mark can be awarded as the conclusion is correct for their values. **B1M0A0A0M1B0A1**

In part (b) they have not said what data refers to. **B0**

Student Response C

a) $H_0: \sigma_A^2 = \sigma_B^2$ $\qquad$ $H_1: \sigma_A^2 \neq \sigma_B^2$

$$S_A^2 = \frac{1}{n_A-1}\left(\Sigma X_A^2 - n\bar{X}_A^2\right)$$
$$= \frac{1}{24}\left(1637.37 - 25\times\left(\frac{194.7}{25}\right)^2\right)$$
$$= 5.0436$$

$$S_B^2 = \frac{1}{n_B-1}\left(\Sigma X_B^2 - n\bar{X}_B^2\right)$$
$$= \frac{1}{25}\left(2031.19 - 26\times\left(\frac{227.5}{26}\right)^2\right)$$
$$= 1.6226$$

$$\frac{S_A^2}{S_B^2} = \frac{5.0436}{1.6226} = 3.11$$

$F_{24,25}(0.05) = 1.96$ $\quad$ $3.11 > 1.96$ $\quad$ ∴ Reject H_0

∴ There is sufficient evidence to suggest the variances of the yields of two varieties of patatoes are not same.

b) The samples are from normally distributed population.
The samples are independent with each other

7/8

Examiner Comments

In part (a) a fully correct solution is given. **B1M1A1M1B1A1ft**

In part (b) whilst the idea that something is normally distributed is present, it does not say what is normally distributed. **B0**

Exemplar Question 4

back to Contents Page

4 The continuous random variable X has cumulative distribution function given by

$$F(x) = \begin{cases} 0 & x \leq 0 \\ k\left(x^3 - \frac{3}{8}x^4\right) & 0 < x \leq 2 \\ 1 & x > 2 \end{cases}$$

where k is a constant.

(*a*) Show that $k = \frac{1}{2}$

(1)

(*b*) Showing your working clearly, use calculus to find

(i) E(X)

(ii) the mode of X

(6)

(*c*) Describe, giving a reason, the skewness of the distribution of X

(1)

(Total for Question 4 is 8 marks)

Mean Score 6.8 out of 8

Examiner Comments

This question is assessing the student's understanding of the relationship between the probability density function and the cumulative distribution function in order to find the mean and mode. (2.1), (2.2) and (2.3).

Part (a) requires working to be shown as the answer is given. Stating $2k = 1$ is not enough as this could easily be gained from the answer. There needed to be clear evidence that 2 had been substituted into *F(x)*.

In part (b) the students had to realise that they needed to differentiate $F(x)$ to gain $f(x)$ before being able to find *F(X)*. They also then needed to differentiate for a second time in order to find the mode. The question asked for clear working to be shown so using a calculator to do the integration was not accepted. It is worth noting that integrating *x F(x)* also gives the answer 1.2 so working must be checked carefully.

In part (c) the students were expected to describe the skewness of the distribution and use their answers for the mode and mean to give a reason, although a correct sketch was also acceptable.

Mark Scheme

Question	Scheme	Marks	AOs
4(a)	$k\left(2^3-\frac{3}{8}2^4\right)=1$	B1*	1.1b
	$2k=1$		
	$k=\frac{1}{2}*$		
	Or $\frac{1}{2}\left(2^3-\frac{3}{8}2^4\right)=1\quad\therefore k=\frac{1}{2}*$	(B1*)	
		(1)	
(b)	$\mathrm{f}(x)=k\left[3x^2-\frac{3}{2}x^3\right]$	M1	2.1
	(i) $\int_0^2 x\mathrm{f}(x)\,\mathrm{d}x=k\int_0^2\left(3x^3-\frac{3}{2}x^4\right)\mathrm{d}x$	M1d	1.1b
	$=\left[\frac{3x^4}{8}-\frac{3x^5}{20}\right]_0^2$		
	$=\frac{6}{5}$ or 1.2	A1	1.1b
	(ii) $3x-\frac{9x^2}{4}=0$	M1d	3.1a
	$x\left(3-\frac{9x}{4}\right)=0$	M1d	1.1b
	$x=0$ or $\frac{4}{3}\ \therefore$ mode $=\frac{4}{3}$	A1	1.1b
		(6)	
(c)	Mode > mean implies it is negative skew	B1ft	2.4
		(1)	
			(8 marks)

Notes:		
(a)	**B1***	substituting x = 2 into F(x) and equating to 1 leading to $k = \frac{1}{2}$ with no errors. Minimum subst seen is k(8 – 6) = 1 or $0.5(8-6)=1$
(b)	**M1**	Realising they need to find the pdf and attempting to differentiate $k\left[x^3 - \frac{3}{8}x^4\right]$ at least 1 correct term
(i)	**M1d**	dep on 1st M1 Attempting tofind $\int_0^2 x(\text{their f}(x))\,dx$ At least one correct term ft their pdf
	A1	$\frac{6}{5}$ or 1.2 oe NB 1.2 with no working gains M0M0A0
(ii)	**M1d**	dep on 1st M1 for realising they need to differentiate their pdf. At least one correct term but ft their pdf
	M1d	Dep on 3rd M1. correct method for solving their differential of their pdf = 0pdf must be of the form ax^2+bx
	A1	$\therefore$ mode $= \frac{4}{3}$ only. They must eliminate 0
(c)	**B1ft**	ft their mode and mean or a correct sketch.

Student Response A

a) $k\left[x^3 - \frac{3}{6}x^4\right]_0^2 = 1$

$2k = 1$

$k = \frac{1}{2}$

b) i) $\frac{1}{2}\left(3x^2 - \frac{3}{2}x^3\right)$

ii) no, 2

$E(X) = \frac{1}{2}\int_0^2 3x^3 - \frac{3}{2}x^4$

$= \frac{1}{2}\left[\frac{3x^4}{4} - \frac{3}{10}x^5\right]_0^2$

$= \frac{1}{2}(12 - 9.6)$

$= 1.2$

c) negativaly skewed.

3/8

Examiner Comments

In part (a) the minimum evidence required for showing 2 has been substituted is not seen. **B0**

In part (b) the pdf has been found. This has then been integrated and the correct limits used to find *E(X)*. There is no correct method given for the mode. **M1M1dA1M0M0A0**

In part (c) no reason is given. **B0**

Student Response B

a) $F(2)=1$

$k\left(2^3 - \frac{3}{8} \times 2^4\right) = 1$

$k(8 - 3 \times 2) = 1$

$2k = 1$

$k = \frac{1}{2}$

b) i) $F(x) = \frac{1}{2}\left(x^3 - \frac{3}{8}x^4\right)$

$f(x) = F'(x) = \frac{1}{2}\left(3x^2 - \frac{3}{4}x^3\right)$

$E(X) = \int_0^2 x \cdot \frac{1}{2}\left(3x^2 - \frac{3}{4}x^3\right) dx$

$= \int_0^2 \frac{1}{2}\left(3x^3 - \frac{3}{4}x^4\right) dx$

$= \frac{1}{2}\left[\frac{3x^4}{4} \quad \frac{3x^5}{4 \times 5}\right]_0^2$

$= \frac{1}{2}\left[\frac{3 \times 2^4}{4} - \frac{3 \times 2^5}{4 \times 5} - 0\right]$

$= \frac{18}{5}$

ii) $f'(x) = \frac{1}{2}\left(6x - \frac{9}{4}x^2\right)$

$f'(x) = 0$

$6x - \frac{9}{4}x^2 = 0$

$x\left(6 - \frac{9}{4}x\right) = 0$

$x_1 = 0$

~~x_2~~ $6 = \frac{9}{4}x$

$x = \frac{8}{3}$

$\therefore x = \frac{8}{3}$ only

c) Mean $= 3.6$

Mode $= 2.66\ldots$

~~Median : $F(x) = \frac{1}{2}$~~

~~$\frac{1}{2}\left(x^3 - \frac{3}{8}x^4\right) = \frac{1}{2}$~~

Mode $(2.66\ldots)$ < Median < Mean (3.6)

~~$x^3 - \frac{3}{8}x^4 = 1$~~

~~$x^3\left(1 - \frac{3}{8}x\right) = 1$~~

Hence positive skewness

6/8

Examiner Comments

In part (a) the minimum evidence required for showing 2 has been substituted is seen. **B1**

In part (b) an attempt to find the pdf has been made and one of the terms is correct. They have then used this pdf and attempted to find *E(X)* the pdf has been found. This has then been integrated correctly and the correct limits used to find *E(X)*. However, the answer is incorrect. **M1M1dA0**

Their pdf has been differentiated correctly, been equated to zero and solved but the answer is incorrect. **M1M1dA0**

In part (c) the correct conclusion using their mean and mode has been made along with a relevant reason. **B1**

Student Response C

(a) $F(2) = 1$ $\qquad k(8 - \frac{3}{8} \times 2^4) = 1$

$k(8 - 3 \times 2) = 1$

$k \times 2 = 1$

$k = \frac{1}{2}$

(b) (i) $F(x) = \frac{1}{2}x^3 - \frac{3}{16}x^4$ for $x \in (0, 2]$

$f(x) = \frac{dF(x)}{dx} = \frac{3}{2}x^2 - \frac{3}{4}x^3$

$E(X) = \int_0^2 x f(x) = \int_0^2 \frac{3}{2}x^3 - \frac{3}{4}x^4 dx$

$= [\frac{3}{8}x^4 - \frac{3}{20}x^5]_0^2$

$= 1.2$

(ii) $f'(x) = \cancel{\frac{3}{4}x} \; 3x - \frac{9}{4}x^2 = 0$

$x = 0,$ $\qquad \frac{9}{4}x = 3$

$x = \frac{4}{3}$

$\therefore x = \frac{4}{3}$ is the mode.

(c) median < mean < mode. it's a negative skew.

$1 < 1.2 < 1.3\dot{3}$

8/8

Examiner Comments

In part (a) the minimum evidence required for showing 2 has been substituted is seen. **B1**

In part (b) the pdf has been found. This has then been integrated and the correct limits used to find *E(X)*. The pdf has been differentiated correctly, been equated to zero and solved. **M1M1dA1M1M1A1**

In part (c) they have the mode > mean leading to the correct conclusion. We can ignore the reference to the incorrect median. **B1**

Exemplar Question 5

back to Contents Page

5 Alexa believes that students are equally likely to achieve the same percentage score on each of two tests, paper I and paper II. She randomly selects 8 students and gives them each paper I and paper II. The percentage scores for each paper are recorded.
The following paired data are collected.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Paper I (%)	70	70	84	80	64	65	65	90
Paper II (%)	64	76	72	74	68	64	58	76

Test, at the 1% significance level, whether or not there is evidence to support Alexa's belief.
State your hypotheses clearly and show your working.

(7)

(Total for Question 5 is 7 marks)

Mean Score 5.1 out of 7

Examiner Comments

This question is assessing the ability to apply a paired *t*-test. (7.2)

The most common error was to use a difference of means test.

Mark Scheme

Question	Scheme	Marks	AOs
5	d: 6 –6 12 6 –4 1 7 14	M1	3.1b
	$\bar{d} = \pm 4.5$ $s_d = \sqrt{50.285...} = 7.09 \ldots$	M1	1.1b
	$H_0 : \mu_d = 0$ $H_1 : \mu_d \neq 0$	B1	3.3
	$t = \pm \dfrac{"4.5"\sqrt{8}}{"7.09..."}$ oe	M1	1.1b
	$= \pm\ 1.7948 \ldots$ awrt $\pm\ 1.79/1.8$	A1	1.1b
	Critical value $t_7 = \pm\ 3.499$	B1	1.1b
	There is insufficient evidence that the **papers** are of a different level of difficulty or **Alexa's** belief is correct	A1ft	2.2b
		(7)	

(7 marks)

Notes:

M1: for realising that the model to use is the paired t-test and finding the differences $(\pm)$ At least 3 correct

M1: correct method for finding $\bar{d}$ and s_d .

B1: Using a correct model for difference and both hypotheses correct using the notation μ_d or μ

Condone $\mu_I = \mu_{II}$ and $\mu_I \neq \mu_{II}$

M1: Using the correct method to find test statistics ie $t = \pm \dfrac{"\text{their } 4.5"\sqrt{8}}{"\text{their } 7.09..."}$

A1: awrt1.79 or 1.8

B1: for correct critical value $t = \pm\ 3.499$ with compatible sign

A1ft: Drawing a correct inference in context using their CV and their value of t

NB difference of means test gets M0M0B1M0A0B0A0

Student Response A

Paper 1	Paper 2
$\sum x = 588$	$\sum x = 552$
$\sum x^2 = 43902$	$\sum x^2 = 38392$
$\bar{x} = 73.5$	$\bar{x} = 69$
$s^2 = \frac{43902 - \left(\frac{588}{8}\right)^2 \times 8}{7}$	$s^2 = \frac{38392 - 69^2(8)}{7}$
$= \frac{684}{7}$	$= \frac{304}{7}$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

two tailed 1% significance

$\therefore$ 0.5% each tail.

CR: t_{14} (0.5%) = 2.977

$$S_p^2 = \frac{(7)\left(\frac{684}{7}\right) + (7)\left(\frac{304}{7}\right)}{14}$$

$$= \frac{494}{7}$$

test stat: $\frac{(73.5 - 69) - (0)}{\sqrt{\frac{494}{7}\left(\frac{1}{8} + \frac{1}{8}\right)}}$

$$= \frac{4.5}{4.2003...}$$

$$= 1.071341813$$

$1.071341813 < 2.977$

$\therefore$ accept Ho, there is not enough ~~s~~ sufficient evidence to suggest percentage scores are different.

Alexa's belief was correct.

1/7

Examiner Comments

A difference in means test has been attempted rather than a paired *t*-test. The only mark they could get is the first B1. They have the relevant hypotheses. **M0M0B1M0A0B0A0**

Student Response B

Paired sample t-test.

Let D be the difference between ~~P~~ Paper I & Paper II

	Student	A	B	C	D	E	F	G	H
~~I−II~~.	I − II(/)	6	−6	12	6	−4	1	7	14.

Hypothesis : H_0 : $D = 0$.

H_1 : $D \neq 0$ at 1% level of significe.

~~sum~~ $\mu = \frac{9}{2}$. $\sigma^2 = \Sigma d^2 - \frac{(\Sigma d)^2}{n}$

$= 514 - \frac{(26)^2}{8} = 352.$

Test statistic : $\frac{\bar{D} - \mu}{\frac{\sigma^2}{\sqrt{n}}} \sim t_{8-1} = t_7$

$$\left| \frac{0 - (\frac{9}{2})}{\sqrt{\frac{352}{8}}} \right| = |-0.6784| = 0.6784$$

critical value : since its 2-tail test.

c.v. = 3.499.

since test-statistic < critical value

∴ Reject Ho at 1% level of significance

∴ There's insufficient evidence to support Alexa's belief.

3/7

Examiner Comments

A paired *t*-test has been used. The differences have been found. s^2 has not been found. The hypotheses do not contain μ. The correct method using what they believe is s (we condone the use of σ as labelling) but the answer is incorrect. The critical value is correct. The conclusion is incorrect as they say there is insufficient evidence to support Alexa's belief.

M1M0B0M1A0B1A0

Student Response C

$H_0: \mu_d = 0$

$H_1: \mu_d \neq 0$

6

~~4~~ -6

12

6

-4

1

7

14

$\bar{x}_d = 4.5$

$\Sigma x^2 = 514$

$$s^2 = \frac{514 - 8 \times 4.5^2}{7}$$

$$= 50.29$$

$$\frac{4.5 - 0}{\sqrt{\frac{50.29}{8}}}$$

$= 1.795 < 2.998$

$\therefore$ insufficient evidence to reject H_0.

$\therefore$ Alexa is correct.

6/7

Examiner Comments

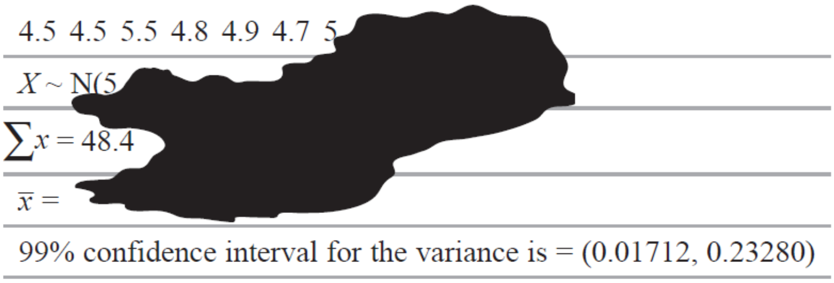
A paired *t*-test has been used. The hypotheses and test statistic are correct but the critical value is incorrect. The conclusion is correct using their CV. **M1M1B1M1A1B0A1**

back to Contents Page

Exemplar Question 6

6 A company manufactures bolts. The diameter of the bolts follows a normal distribution with a mean diameter of 5 mm.

Stan believes that the mean diameter of the bolts is less than 5 mm. He takes a random sample of 10 bolts and measures their diameters. He calculates some statistics but spills ink on his work before completing them. The only information he has left is as follows



4.5 4.5 5.5 4.8 4.9 4.7 5

$X \sim \mathrm{N}(5$

$\sum x = 48.4$

$\bar{x} =$

99% confidence interval for the variance is = (0.01712, 0.23280)

Stating your hypotheses clearly, test, at the 5% level of significance, whether or not Stan's belief is supported.

(9)

(Total for Question 6 is 9 marks)

Mean Score 6.8 out of 9

Examiner Comments

This question is assessing material in (5.2, 6.1 and 7.2) and is an example of a problem solving question where the students need to devise their own strategy for solving it.

Students generally knew they needed to use for the mean of a normal distribution but struggled to find a value forthe unbiased estimator of σ. The most common error was to think that the confidence interval was in fact for the mean rather than the variance.

Mark Scheme

Question	Scheme	Marks	AOs
6	99% confidence interval for Var uses χ^2 values of 1.735 or 23.589	B1	3.3
	$\frac{9s^2}{1.735} = 0.2328$ or $\frac{9s^2}{23.589} = 0.01712$	M1	2.1
	$s^2 = \frac{0.2328 \times "1.735"}{9}$ or $\frac{0.01712 \times "23.589"}{9}$ $[= 0.04487...]$	dM1	1.1b
	$\bar{x} = 4.84$	B1	1.1b
	$\mathrm{H}_0 : \mu = 5$ $\mathrm{H}_1 : \mu < 5$	B1	2.5
	CV $t_9 = -1.833$	B1	1.1b
	$t = \pm \frac{"4.84" - 5}{\sqrt{"0.0449"/10}}$	M1	1.1b
	= awrt -2.39	A1	1.1b
	Stan's belief is supported **or** there is evidence that the **mean diameter** of the bolts is less than **5mm**	A1ft	2.2b
		(9)	
			(9 marks)

Notes:

B1: For realising a χ^2 distribution must be used as a model and finding a correct value

M1: For realising the need to set $\frac{9s^2}{\text{"smallest } \chi^2\text{"}} = 0.2328$ or $\frac{9s^2}{\text{"largest } \chi^2\text{"}} = 0.01712$

dM1: correct method used to solve equation to find s^2

B1: awrt 4.84

B1: Both hypotheses correctusing the notation μ

B1: ± 1.833

M1: For us of correct formula ie $\pm \frac{\text{"their 4.84"} - 5}{\sqrt{\text{"their 0.0449"}/10}}$ If "4.84" not shown it must be correct here

A1: – 2.39

A1ft: Drawing a correct inference following through their CV and test statistic (must have matching signs)

NB if chi squared values not shown

$s^2 = 0.045$ or 0.0449 award B0 M1M1 forawrt 0.04487 award B1 M1 A1

Use of $2(2.5758)\frac{\sigma}{\sqrt{10}} = 0.21568$ gives $\sigma = \sqrt{0.0175}$ could get B0M0M0B1B1B1M0A0A0

Unless continue to get $s^2 = \frac{10}{9}0.0175 = 0.0194...$

Use of $2(1.833)\frac{s}{\sqrt{10}} = 0.21568$ gives $s = 0.1860$ could get B0M0M0B1B1B1M1A0A1

Student Response A

$H_0 : \mu = 5 \text{ mm}$ $\quad$ $H_1 : \mu < 5 \text{ mm}$

$X \sim N(5, \sigma^2)$

normal distribution so $\sigma^2 = \dfrac{0.2328 + 0.01712}{2} = 0.12496$

$\sigma = 0.3535$

$\therefore X \sim N(5, 0.12496)$

~~$X \sim N(5, 0.12496$~~

$\bar{X} \sim N\left(5, \left(\dfrac{0.3535}{\sqrt{10}}\right)^2\right)$

$\bar{x} = \dfrac{48.4}{10} = 4.84$

$\therefore P(\bar{X} \leq 4.84) = 0.0762$

$0.0762 > 0.05$

$\therefore$ H_0 should be accepted and it can be concluded that based on these data there is insufficient evidence to suggest the mean diameter of the bolts is less than 5mm.

2/9

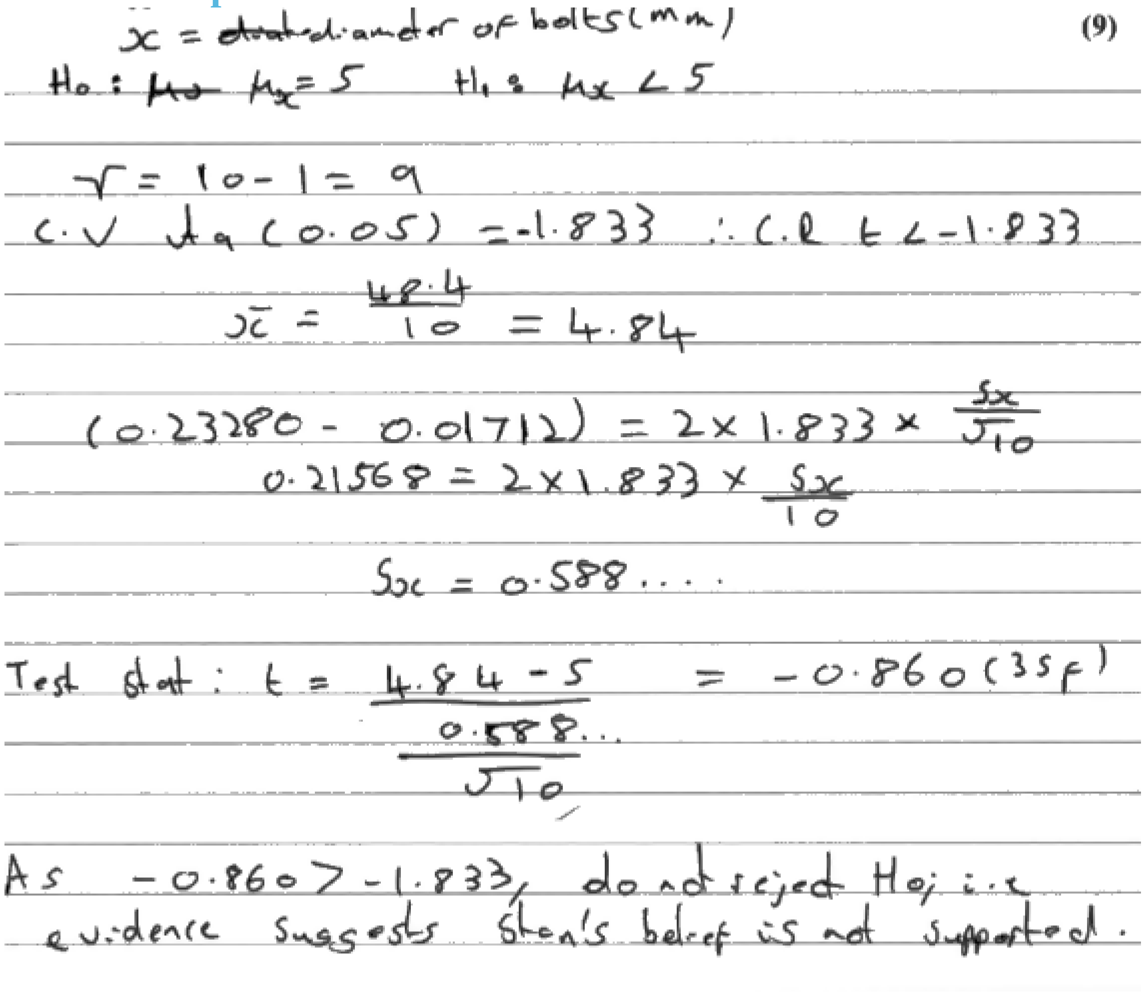
Examiner Comments

The student has not used the confidence interval for the variance to find s^2. The hypotheses and mean are correct. The method used for the test statistic is not shown and the value is incorrect. No critical value is given. The final mark for the conclusion cannot be awarded as the previous method mark is not awarded.

B0M0dM0B1B1B0M0A0A0ft

Student Response B

id 0461004630630



(9)

x = ~~[illegible]~~ diameter of bolts (mm)

H_0: ~~μ_0~~ $\mu_x = 5$ $\quad H_1$: $\mu_x < 5$

$\nu = 10 - 1 = 9$

c.v $t_9(0.05) = -1.833$ $\therefore$ C.R $t < -1.833$

$\bar{x} = \frac{48.4}{10} = 4.84$

$(0.23280 - 0.01712) = 2 \times 1.833 \times \frac{s_x}{\sqrt{10}}$

$0.21568 = 2 \times 1.833 \times \frac{s_x}{10}$

$s_x = 0.588 \ldots$

Test stat: $t = \dfrac{4.84 - 5}{\frac{0.588\ldots}{\sqrt{10}}} = -0.860$ (3sf)

As $-0.860 > -1.833$, do not reject H_0; i.e. evidence suggests Stan's belief is not supported.

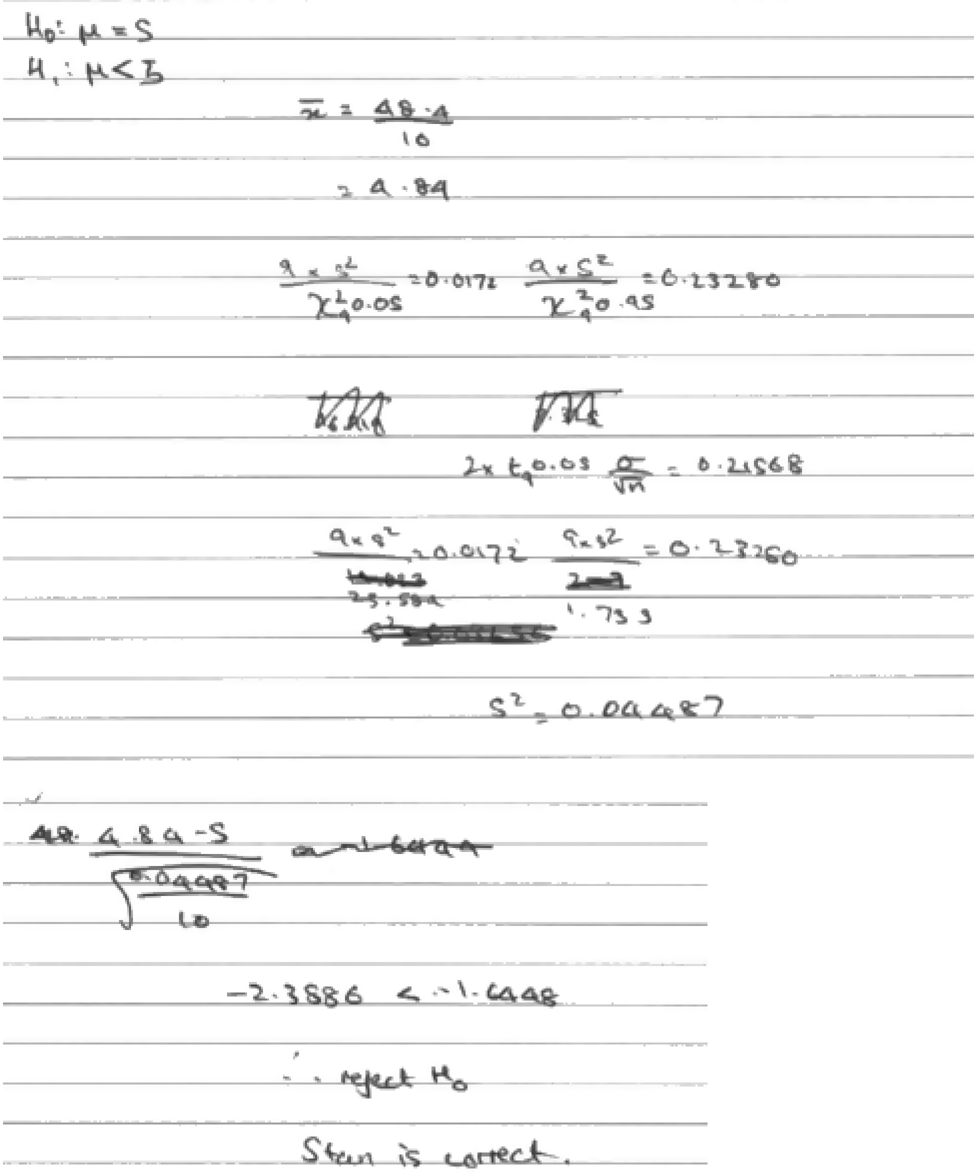
5/9

Examiner Comments

The student has not used the confidence interval for the variance to find s^2. The hypotheses, mean and critical value are correct. The method used for the test statistic is correct using their *s* but the answer is incorrect. The conclusion is correct using their test statistic and contains the word Stan.

B0M0dM0B1B1B1M1A0A1ft

Student Response C



$H_0: \mu = 5$

$H_1: \mu < 5$

$$\bar{x} = \frac{48.4}{10}$$

$$= 4.84$$

$$\frac{9 \times s^2}{\chi^2_9 0.05} = 0.0172 \qquad \frac{9 \times s^2}{\chi^2_9 0.95} = 0.23280$$

$$2 \times t_9 0.05 \frac{\sigma}{\sqrt{n}} = 0.21568$$

$$\frac{9 \times s^2}{23.584} = 0.0172 \qquad \frac{9 \times s^2}{1.733} = 0.23260$$

$$s^2 = 0.04487$$

$$\frac{4.84 - 5}{\sqrt{\frac{0.04487}{10}}} \quad \cancel{= -1.6444}$$

$$-2.3886 < -1.6448$$

$\therefore$ reject H_0

Stan is correct.

8/9

Examiner Comments

The student has found the correct value for s^2. The hypotheses and test statistic are correct, but the critical value is incorrect. The conclusion is correct and contains the word Stan.

B1M1dM1B1B1B0M1A1A1ft

Exemplar Question 7

back to Contents Page

7 A manufacturer makes two versions of a toy. One version is made out of wood and the other is made out of plastic.

The weights, W kg, of the wooden toys are normally distributed with mean 2.5 kg and standard deviation 0.7 kg. The weights, kg, of the plastic toys are normally distributed with mean 1.27 kg and standard deviation 0.4 kg. The random variables W and X are independent.

(*a*) Find the probability that the weight of a randomly chosen wooden toy is more than double the weight of a randomly chosen plastic toy.

(6)

The manufacturer packs n of these wooden toys and $2n$ of these plastic toys into the same container. The maximum weight the container can hold is 252 kg.

The probability of the contents of this container being overweight is 0.2119 to 4 decimal places.

(*b*) Calculate the value of n.

(8)

(Total for Question 7 is 14 marks)

Mean Score 9.5 out of 14

Examiner Comments

This question is assessing student's understanding of testing for the mean of a Poisson distribution (4.1).

In part (a) students need to realise that they need to use $T = W - 2X$ and therefore $\mathrm{E}(T) = \mathrm{E}(W) - 2\mathrm{E}(X)$ and $\mathrm{Var}(T) = \mathrm{Var}(W) - 2^2\mathrm{Var}(X)$.

In part (b) students need to realise that they need to use $B = W_1 + W_2 + \ldots + W_n + X_1 + X_2 + \ldots + X_{2n}$ and therefore $\mathrm{E}(B) = \mathrm{E}(W) - 2\mathrm{E}(X)$ and $\mathrm{Var}(T) = \mathrm{Var}(W) - 2^2\mathrm{Var}(X)$. The most common error was to use the distribution$B = 2X + 2nW$.

Mark Scheme

Question	Scheme	Marks	AOs
7(a)	Let $T = W - 2X$ then $\mathrm{E}(T) = 2.5 - 2\times 1.27$	M1	3.3
	$= -0.04$	A1	1.1b
	$\mathrm{Var}(T) = 0.7^2 + 2^2 \times 0.4^2$	M1	2.1
	$= 1.13$	A1	1.1b
	$\mathrm{P}\left(Z > \frac{0 - "-0.04"}{\sqrt{"1.13"}}\right) = \mathrm{P}(Z > 0.0376\ldots)$	M1	2.1
	= awrt 0.484/0.485	A1	1.1b
		(6)	
(b)	$B = W_1 + W_2 + \ldots + W_n + X_1 + X_2 + \ldots + X_{2n}$	M1	3.3
	$\mathrm{E}(B) = 5.04n$	B1	1.1b
	$\mathrm{Var}(B) = n\times 0.7^2 + 2n \times 0.4^2$		
	$= 0.81n$	A1	1.1b
	$\pm\frac{252 - "5.04n"}{\sqrt{"0.81n"}}$	M1	1.1b
	$\frac{252 - "5.04n"}{\sqrt{"0.81n"}} = 0.8$	M1	2.1
	$5.04n + 0.72\sqrt{n} - 252 = 0$ oe		
	$\sqrt{n} = -7.14\ldots$ or 7	M1	1.1b
	$n = 7^2$	M1	1.1b
	$= 49$	A1cso	1.1b
		(8)	

(14 marks)

Notes:

(a) M1: selecting and using an appropriate model. ie $\pm(W - 2X)$ May be implied by -0.04

A1: -0.04 oe

M1: for realising the need to use Var(W) + 4 Var(X). Allow use of 0.7 for Var (W) instead of 0.7^2 and/or 0.4 for Var(X) instead of 0.4^2. May be implied by 1.13

A1: 1.13 only

M1: For realising the P(T> 0) is required and an attempt to find it. $\frac{0 - "\text{their } -0.04"}{\sqrt{"\text{their } 1.13"}}$ may be implied by a correct answer. If E(T) and Var(T) have not been given they must be correct here

A1: awrt 0.484/0.485

(b)M1: Selecting and using appropriate model. May be implied by 0.81

B1: 5.04n only

A1: 0.81n

M1: For standardising using their mean and sd $\pm\frac{252 - "5.04n"}{\sqrt{"0.81n"}}$ If mean and sd not given they must be correct here

M1: For constructing an equation and equate their standardisation to 0.8 or awrt 0.7998. Must be of form $\frac{252 - an}{b\sqrt{n}} = 0.8$ or $\frac{252 - an}{bn} = 0.8$

M1: Correctly solving their 3 term quadratic equation. Condone $n = 7$

M1: for realising the need to square their answer or for attempting to square their quadratic equation

A1cso: 49 only

Student Response A

a) $W \sim N(2.5, 0.7)$
$X \sim N(1.27, 0.4)$

let $Y = W - 2X$

$Y \sim N(-0.04, 2.3)$

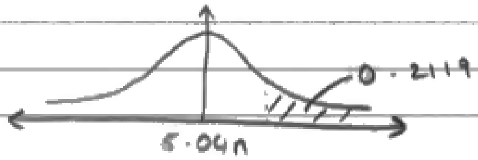
$\mu = 2.5 - 2 \times 1.27 = -0.04$
$\text{Var}(Y) = 0.7 + 4 \times 0.4 =$

if wooden toy is over double plastic toy in weight,
$Y > 0$

$P(Y > 0) = 0.4895$

b) let $Z = nW + 2nX$
$Z \sim N(5.04n, 2.3n^2)$

$P(Z > 252) = 0.2119$



$$z = \frac{x - \mu}{\theta}$$

$$0.7998 = \frac{252 - 5.04n}{2.3n^2}$$

$1.8396n^2 + 5.04n - 252 = 0$

$n = 10.41$ or $n = -13.15$
(not valid)

$\downarrow$

$\therefore n = 10$

3/14

Examiner Comments

In part (a) E($W - 2X$) is correct but the method used to find Var($W - 2X$) is incorrect. **M1A1M0A0**
The standardisation is not seen, and the answer is incorrect. **M0A0**

In part (b) E(B) is correct but Var(B) is incorrect as the incorrect distribution $B = 2X + 2nW$ has been used. **M0B1A0**
The standardisation is incorrect as they have used their variance rather than their standard deviation. **M0**
Although they have equated their standardisation to 0.7998 it is not one of the required forms. **M0**
No working has been shown when solving the quadratic equation, so we do not know if the equation has been solved correctly. **M0**
Since they used the variance in their standardisation, they do not square anything. **M0A0**

Student Response B

a) $p(W>2X) = p(W-2X>0)$

$W-2X \sim N(\mu_W - 2\mu_X, \sigma_W^2 + 4\sigma_X^2)$

$W-2X \sim N(-0.04, 1.13)$

$p(W-2X>0)$

$= \Phi\left(\frac{-0.04}{\sqrt{1.13}}\right) = \Phi(-0.03763)$

$= 0.4850$

b) $nW + 2nX \sim N(n(\mu_W + 2\mu_X), n^2\sigma_W^2 + 4n^2\sigma_X^2)$

$p(nW+2nX > 252)$

$= \Phi\left(\frac{252-5.04n}{n\sqrt{\sigma_W^2+4\sigma_X^2}}\right) = \Phi\left(\frac{252-5.04n}{n\sqrt{1.13}}\right) = \cancel{0.2119}$ 0.8 7881

$\frac{252-5.04n}{n\sqrt{1.13}} = \Phi^{-1}(0.2119) = +0.800$

$\Rightarrow 252 - 5.04n = +0.8n\sqrt{1.13}$ $\Rightarrow 252 = n(5.04 + 0.8\sqrt{1.13})$

$\Rightarrow$ [illegible]

~~$n = 60.15$~~

$n = \cancel{60.}43$

8/14

Examiner Comments

In part (a) the student has the correct answer from correct working **M1A1M1A1M1A1**

In part (b) E(*B*) is correct but Var(*B*) is incorrect in the standardisation as the incorrect distribution *B* = *nW* + 2*nX* has been used. **M0B1A0**
No values for E(*B*) and Var(*B*) have been given so they must be correct in the standardisation to gain the method mark. They are incorrect. **M0**
Their standardisation has been equated to 0.8. **M1**
They have not got a quadratic equation to solve. **M0**
Neither the equation or the answer have been squared. M**0A0**

Student Response C

(8)

a) $W \sim N(2.5, 0.7^2)$ $X \sim N(1.27, 0.4^2)$

$W > 2X$

$W - 2X > 0$

$W-2X \sim N(2.5 - 2(1.27), 0.7^2 + 4(0.4^2))$

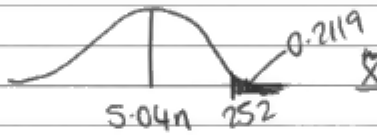
$W-2X \sim N(-0.04, 1.13)$

$P(W-2X > 0) = 0.485$

b) $nW + 2nX \sim N(2.5n + 2.54n, 0.7^2 n + 0.4^2 \times 2n)$

$\sim N(5.04n, 0.49n + 0.32n)$

$nW + 2nX \sim N(5.04n, 0.81n)$



$\frac{x-\mu}{\sigma} = \frac{252 - 5.04n}{\sqrt{0.81n}} = 0.2119$

$252 - 5.04n = 0.2119 \times 0.9\sqrt{n}$

$252 - 5.04n - 0.19071\sqrt{n} = 0$

$5.04n + 0.19071\sqrt{n} - 252 = 0$

let $\sqrt{n} = x$

$5.04x^2 + 0.19071x - 252 = 0$

$x = 7.05, -7.09$

$\sqrt{n} = x$

$\sqrt{n} = 7.05$

$n = 49.7025$

$n = 49$

11/14

Examiner Comments

In part (a) the student has the correct answer from correct working **M1A1M1A1M1A1**

In part (b) E(*B*) and Var(*B*) and the standardisation are correct. **M1B1A1M1**
The standardisation has been equated to 0.2119 instead of 0.8. **M0**
No working has been shown when solving the quadratic equation, so we do not know if the equation has been solved correctly. **M0**
They have realised that they need to square the answer. **M1**
Although 49 is gained it has come from incorrect working so this is not a correct solution only. **A0**

Exemplar Question 8

back to Contents Page

8 Nine athletes, *A*, *B*, *C*, *D*, *E*, *F*, *G*, *H* and *I*, competed in both the 100 m sprint and the long jump. After the two events the positions of each athlete were recorded and Spearman's rank correlation coefficient was calculated and found to be 0.85

(*a*) Stating your hypotheses clearly, test whether or not there is evidence to suggest that the higher an athlete's position is in the 100 m sprint, the higher their position is in the long jump. Use a 5% level of significance.

(4)

The piece of paper the positions were recorded on was mislaid. Although some of the athletes agreed their positions, there was some disagreement between athletes *B*, *C* and *D* over their long jump results.

The table shows the results that are agreed to be correct.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Position in 100 m sprint	4	6	7	9	2	8	3	1	5
Position in long jump	5				4	9	3	1	2

Given that there were no tied ranks,

(*b*) find the correct positions of athletes *B*, *C* and *D* in the long jump. You must show your working clearly and give reasons for your answers.

(5)

(*c*) Without recalculating the coefficient, explain how Spearman's rank correlation coefficient would change if athlete *H* was disqualified from both the 100 m sprint and the long jump.

(2)

(Total for Question 8 is 11 marks)

Mean Score 7.6 out of 11

Examiner Comments

This question is assessing the student's understanding of Spearman's rank correlation coefficient (3.2). It also tests the students are able to test the hypothesis that a correlation is zero (3.3).

Part (a) is a straightforward test of the hypothesis that a correlation is zero.

Part (b) requires clear working and full explanation to solve a problem. Most students were able to allocate the positions correctly with many showing clear working. Very few students were able to give a satisfactory reason as to why *B* must be in position 7 or alternatively why *D* was in 8th.

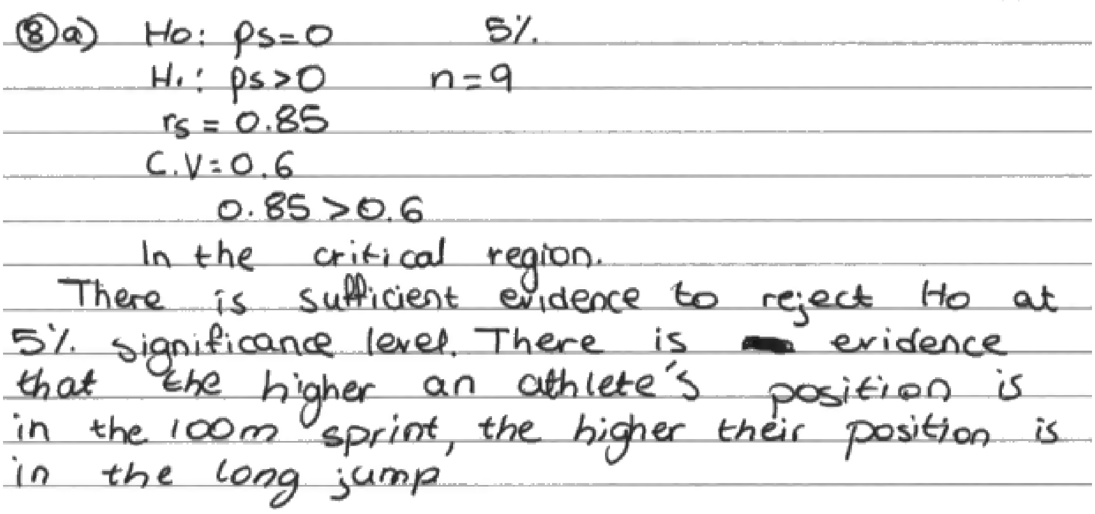
Part (c) requires a full explanation. The most common explanation given was that *d* would not change however this needs to be linked to the sum of d^2 not changing as this is what is used in the formulae.

Mark Scheme

Question	Scheme	Marks	AOs
8(a)	H$_0$: $\rho_s = 0$ H$_1$: $\rho_s > 0$	B1	2.5
	CV = 0.6	B1	1.1b
	r_s = 0.85 does lie in the critical region	M1	2.1
	There is evidence to suggest that there is a relationship between the position in the 100m sprint and the position in the long jump.	A1	2.2b
		(4)	
(b)	$1 - \dfrac{6\sum d^2}{9(80)} = 0.85$	M1	3.1b
	$\sum d^2 = 18$	A1	1.1b
	$\sum d^2$ needed is $'18' - 15 = 3$	M1	1.1b
	Since $\sum d^2 = 3$ for the 3 missing places each place must contribute 1, therefore *B* must be in position 5 or 7. However, 5 has already been used so they must be position 7	A1	2.2a
	C is 6th and *D* is 8th	A1	2.2a
	SC B7, C6, D8 with no reasons B1 marks as final A1 on epen		
		(5)	
(c)	The $\sum d^2$ will not change but the value of *n* will decrease therefore	M1	2.4
	Spearman's rank correlation will decrease	A1	2.2a
		(2)	
	Notes:	**(11 marks)**	

(a)B1: Both hypotheses correct written using the notation ρ
B1: awrt 0.6
M1: Drawing a correct inference using their CV and the value of r_s
A1: Drawing a correct inference in context using their CV and the value of r_s
(b)M1: For realising they need to equate $1 - \dfrac{6\sum d^2}{9(80)}$ to 0.85 to enable them to find the $\sum d^2$
A1: 18
M1: for $\sum d^2$ =3
A1: Forusing the information in the question with the value for $\sum d^2$ to deduce that each must contribute 1 to the $\sum d^2$ and explain why *B* must be in position 7
A1: C 6th D 8th
(c)M1: Complete explanation why it decreases
A1: using the information given to deduce that it decreases

Student Response A



⑧a) $H_0: \rho_s = 0$ 5%

$H_1: \rho_s > 0$ $n = 9$

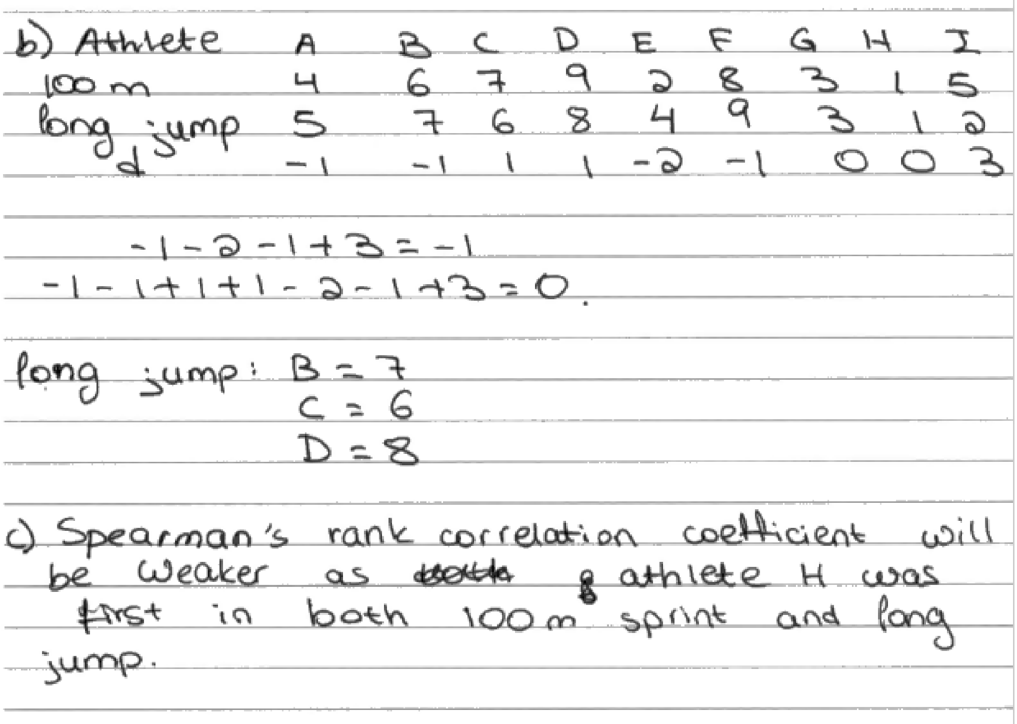
$r_s = 0.85$

C.V $= 0.6$

$0.85 > 0.6$

In the critical region.

There is sufficient evidence to reject Ho at 5% significance level. There is ~~[illegible]~~ evidence that the higher an athlete's position is in the 100m sprint, the higher their position is in the long jump



b) Athlete	A	B	C	D	E	F	G	H	I
100 m	4	6	7	9	2	8	3	1	5
long jump	5	7	6	8	4	9	3	1	2
d	−1	−1	1	1	−2	−1	0	0	3

$-1-2-1+3=-1$

$-1-1+1+1-2-1+3=0.$

long jump: B = 7
C = 6
D = 8

c) Spearman's rank correlation coefficient will be weaker as ~~both~~ athlete H was first in both 100m sprint and long jump.

5/11

Examiner Comments

In part (a) a fully correct solution with the required context in the conclusion is given. **B1B1M1A1**

Part (b) is an example of the special case. The positions are correct but no explanation has been given. **M0A0M0A0A1**

In part (c) the explanation is incomplete. **M0A0**

Student Response B

a) $H_0 : \rho = 0$ $H_1 : \rho > 0$

$0.85 > 0.6$

therefore, ~~we~~ there is sufficient evidence to reject H_0

b) $1 - \frac{6\Sigma d^2}{n(n^2-1)} = 0.85$

$1 - \frac{6\Sigma d^2}{9(9^2-1)} = 0.85$

$\frac{6\Sigma d^2}{9(9^2-1)} = 1 - 0.85$

~~$6\Sigma d^2$ [illegible] $\Sigma d^2 = 16$~~

$6\Sigma d^2 = 108$ $\Sigma d^2 = 18$

	A	B	C	D	E	F	G	H	I
d	1				2	1	0	0	-3
d^2	1				4	1	0	0	9

$18 - (1 + 4 + 1 + 9) = 3$

$\therefore$ Σd^2 for BCD must equal 3

~~B = [illegible]6 C = [illegible] D = 9~~

B = 6 C = 9 D = 7

c) it would be lower as there would be less agreement on the ranks, so it would be closer to 0

6/11

Examiner Comments

In part (a) the hypotheses are correct and in terms of ρ. **B1**
The Critical value is correct and a correct inference is drawn. **B1M1**
The conclusion contains no context. **A0**

In part (b) they have set up the equation and shown the sum of d^2 is 18. **M1A1**

They have shown clearly that the sum of the three d's must = 3 but not realised that each of B,C and D must contribute 1. All 3 positions suggested are incorrect. **M1A0A0**

In part (c) although they state that Spearman's rank correlation will decrease the explanation is not precise enough. **M0A0**

Student Response C

(2)

a) $H_0: P=0$ $H_1: P>0$

Critical value = 0.6000

Critical region $r \geqslant 0.6000$

$0.85 > 0.600$

$\therefore$ in CR

reject H_0

there is sufficient evidence at the 5% significance level to suggest that the higher an athlete's position in the 100m sprint, the higher their position in the long jump

b) $r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 0.85$

$1 - \frac{6\Sigma d^2}{9(80)} = 0.85$

$\frac{6\Sigma d^2}{720} = 0.15$

$6\Sigma d^2 = 108$

$\Sigma d^2 = 18$

	A	E	F	G	H	I
d	−1	−2	−1	0	0	3
d^2	1	4	1	0	0	9

$1+4+1+9 = 15$ So the Σd^2 for B-D is 3 as $18-15=3$

The only way for this to be true is if they were all one rank higher/lower in long jump compared to their 100m

B must have finished 7th in long jump as A finished 5th

DG must have finished 8th as he can't finish 10th (there are only 9 athletes

This leaves C with 6th

B 7th

C 6th

D 8th

c) ~~the 8~~ Σd^2 would still be the same as H had the same rank, however n would decrease so the denominator would be smaller meaning $\frac{6\Sigma d^2}{n(n^2-1)}$ would increase therefore r_s would increase as you are ~~minusing~~ taking away a larger number from 1

10/11

Examiner Comments

In part (a) a fully correct solution with the required context in the conclusion is given. **B1B1M1A1**

In part (b) a fully correct solution with all the required reasons is given. The equation has been set up and the sum of d^2 found. They have realised that the difference must sum to 3 and each must be 1 rank higher or lower than their rank in the 100m sprint. A relevant reason has been given why *B* must finish 7th rather than 6th and the final positions are correct. **M1A1M1A1A1**

In part (c) they have stated that the sum of d^2 will remain the same whilst n increases. However, they then go on to state that r_s would increase. **M1A0**