OCR
Oxford Cambridge and RSA

Friday 16 June 2023 – Afternoon

A Level Further Mathematics B (MEI)

Y432/01 Statistics Minor

Time allowed: 1 hour 15 minutes

9970716358

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for Further Mathematics B (MEI)
- a scientific or graphical calculator

QP

INSTRUCTIONS
- Use black ink. You can use an HB pencil, but only for graphs and diagrams.
- Write your answer to each question in the space provided in the **Printed Answer Booklet**. If you need extra space use the lined pages at the end of the Printed Answer Booklet. The question numbers must be clearly shown.
- Fill in the boxes on the front of the Printed Answer Booklet.
- Answer **all** the questions.
- Where appropriate, your answer should be supported with working. Marks might be given for using a correct method, even if your answer is wrong.
- Give your final answers to a degree of accuracy that is appropriate to the context.
- Do **not** send this Question Paper for marking. Keep it in the centre or recycle it.

INFORMATION
- The total mark for this paper is **60**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **8** pages.

ADVICE
- Read each question carefully before you start your answer.

© OCR 2023 [M/508/5596]
DC (ST/SG) 318818/5

OCR is an exempt Charity

stemup.co.uk

1 A fair spinner has ten sectors, labelled 1, 2, …, 10. In order to start a game, Kofi has to obtain an 8, 9 or 10 on the spinner.

(a) Find the probability that Kofi starts the game on the third spin. **[2]**

(b) Find the probability that Kofi takes at least 5 spins to start the game. **[1]**

(c) Determine the probability that the number of spins required to start the game is within 1 standard deviation of its mean. **[5]**

2 A company manufactures batches of twenty thousand tins which are subsequently filled with fruit. The company tests tins from each batch to make sure that they are strong enough. The test is easy and cheap to carry out, but when a tin has been tested it is no longer suitable for filling with fruit.

(a) (i) Explain why a sample size of 5 tins per batch may not be appropriate in this case. **[1]**

(ii) Explain why a sample size of 1000 tins per batch may not be appropriate in this case. **[1]**

The company tests a sample of 30 tins from each batch.

(b) Explain why it would **not** be sensible for the sample to consist of the final 30 tins produced in a batch. **[1]**

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

© OCR 2023

stemup.co.uk

3 A fair four-sided dice has its faces numbered 0, 1, 2, 3. The dice is rolled three times. The discrete random variable X is the sum of the lowest and highest scores obtained.

(a) Show that $P(X = 1) = \frac{3}{32}$. **[3]**

The table below shows the probability distribution of X.

r	0	1	2	3	4	5	6
$P(X = r)$	$\frac{1}{64}$	$\frac{3}{32}$	$\frac{13}{64}$	$\frac{3}{8}$	$\frac{13}{64}$	$\frac{3}{32}$	$\frac{1}{64}$

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

Find each of the following.

- $E(X)$
- $\text{Var}(X)$ **[4]**

(c) The random variable Y represents the sum of 10 values of X.

(i) State a property of the 10 values of X that would make it possible to deduce the standard deviation of Y. **[1]**

(ii) Given that this property holds, determine the standard deviation of Y. **[2]**

© OCR 2023

stemup.co.uk

4 Eve lives in a narrow lane in the country. She wonders whether the number of vehicles passing her house per minute can be modelled by a Poisson distribution with mean μ. She counts the number of vehicles passing her house over 100 randomly selected one-minute intervals. The results are shown in **Table 4.1**.

Table 4.1

Number of vehicles	0	1	2	3	4	5	6	7	8	9	10	$\geqslant 11$
Frequency	36	33	14	10	4	1	0	0	1	0	1	0

(a) Use the results to find an estimate for μ. **[1]**

The spreadsheet in **Fig. 4.2** shows data for a χ^2 test to assess the goodness of fit of a Poisson model. The sample mean from part **(a)** has been used as an estimate for the population mean. Some of the values in the spreadsheet have been deliberately omitted.

Fig. 4.2

	A	B	C	D	E
1	Number of vehicles	Observed frequency	Poisson probability	Expected frequency	Chi-squared contribution
2	0	36	0.2725	27.2532	2.8073
3	1	33	0.3543	35.4291	
4	2	14			3.5400
5	$\geqslant 3$	17			0.5145
6					

(b) Calculate the missing values in each of the following cells, giving your answers correct to **4** decimal places.

- C4
- D5
- E3

[4]

(c) **In this question you must show detailed reasoning.**

Carry out the χ^2 test at the 5% significance level. **[6]**

(d) Eve checks her data and notices that the two largest numbers of vehicles per minute (8 and 10) occurred when some horses were being ridden along the lane, causing delays to the vehicles. She therefore repeats the analysis, missing out these two items of data. She finds that the value of the χ^2 test statistic is now 4.748. The number of degrees of freedom of the test is unchanged.

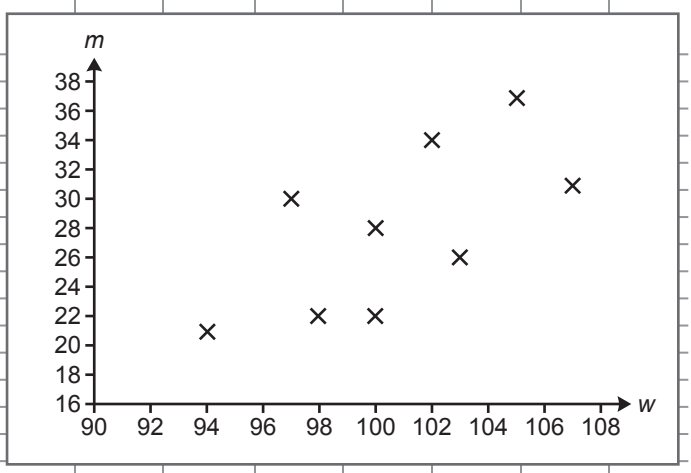
Make **two** comments about this revised test. **[2]**

© OCR 2023

stemup.co.uk

5 An ornithologist is investigating the link between the wing length and the mass of small birds, in order to try to predict the mass from the wing length without having to weigh birds. The ornithologist takes a random sample of 9 birds and measures their wing lengths w mm and their masses m g. The spreadsheet below shows the data, together with a scatter diagram which illustrates the data.

	A	B
1	Wing length w mm	Mass m g
2	100	28
3	98	22
4	107	31
5	100	22
6	97	30
7	103	26
8	105	37
9	102	34
10	94	21



(a) Find the equation of the regression line of m on w, giving the coefficients correct to **3** significant figures. **[2]**

(b) Use the equation which you found in part **(a)** to estimate the mass for each of the following wing lengths.

- 99 mm
- 110 mm

[2]

(c) Comment on the reliability of your estimates. **[2]**

(d) The equation of the regression line of w on m is $w = 0.473m + 87.5$. A friend of the ornithologist suggests that this equation could also be used to estimate the masses of birds from their wing lengths.

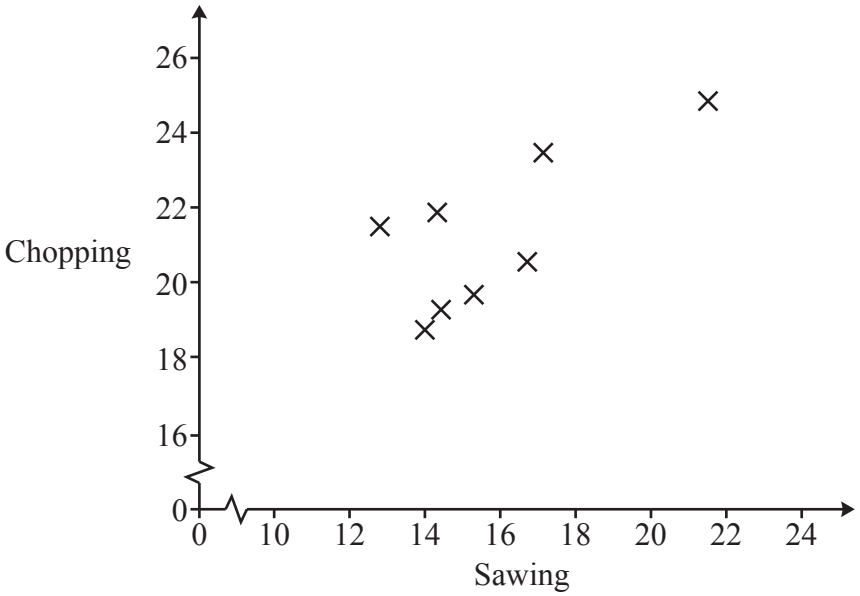
Comment on this suggestion. **[2]**

© OCR 2023

stemup.co.uk

6 Each competitor in a lumberjacking competition has to perform various disciplines for which they are timed. A spectator thinks that the times for two of the disciplines, chopping wood and sawing wood, are related. The table and the scatter diagram below show the times of a random sample of 8 competitors in these two disciplines.

Competitor	A	B	C	D	E	F	G	H
Sawing	17.1	16.7	14.3	14.0	12.8	21.5	15.3	14.4
Chopping	23.5	20.6	21.9	18.8	21.5	24.8	19.7	19.3



(a) The spectator decides to carry out a hypothesis test to investigate whether there is any relationship.

Explain why the spectator decides that a test based on Pearson's product moment correlation coefficient may **not** be valid. **[2]**

(b) Determine the value of Spearman's rank correlation coefficient. **[3]**

(c) Carry out a hypothesis test at the 5% significance level to investigate whether there is positive association between sawing and chopping times. **[5]**

© OCR 2023

stemup.co.uk

7 The discrete random variable X has a uniform distribution over the set of all integers between 100 and n inclusive, where n is a positive integer with $n > 100$.

(a) Given that n is even, determine $\mathrm{P}\left(X < \frac{100+n}{2}\right)$. **[3]**

(b) Determine the variance of the sum of 50 independent values of X, giving your answer in the form $a(n^2 + bn + c)$, where a, b and c are constants. **[3]**

END OF QUESTION PAPER

© OCR 2023

stemup.co.uk

OCR
Oxford Cambridge and RSA

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact The OCR Copyright Team, The Triangle Building, Shaftesbury Road, Cambridge CB2 8EA.

OCR is part of Cambridge University Press & Assessment, which is itself a department of the University of Cambridge.

© OCR 2023

stemup.co.uk

Surname	StemUp Model Solution
Answer booklet	Write the question number in the left-hand margin at the start of each answer.

Question number

1 (a) Let X = number of spins needed. $P(\text{success}) = \frac{3}{10} = 0.3$

$X \sim \text{Geo}(0.3)$

$P(X = 3) = 0.7^2 \times 0.3 = 0.147$

1 (b) $P(X \geqslant 5) = P(\text{first 4 spins all fail}) = 0.7^4 = 0.2401$

1 (c) $E(X) = \frac{1}{p} = \frac{1}{0.3} = \frac{10}{3}$

$\text{Var}(X) = \frac{1-p}{p^2} = \frac{0.7}{0.3^2} = \frac{70}{9}$

$\text{SD}(X) = \frac{\sqrt{70}}{3} = 2.7888...$

$\frac{10}{3} - 2.7888... = 0.544...$, $\frac{10}{3} + 2.7888... = 6.122...$

$P(0.544... < X < 6.122...) = P(1 \leqslant X \leqslant 6) = P(X \leqslant 6)$

$= 1 - 0.7^6 = 1 - 0.117649 = 0.882$ (3 s.f.)

2 (a)(i) A sample of only 5 tins is unlikely to be representative of the batch, so it may not give useful information about strength.

2 (a)(ii) Testing destroys the tins (they cannot be filled with fruit), so testing 1000 is wasteful; a much smaller sample is adequate.

2 (b) There may have been a fault earlier in the batch which then

Question number

corrected itself, so the final 30 tins would not be random

or representative of the whole batch.

2 (c) The sample should be random.

The sample should be representative of the whole batch (unbiased).

3 (a) $X = 1$ means lowest score 0 and highest score 1, so the

three scores are either $\{0, 0, 1\}$ or $\{0, 1, 1\}$.

$\{0, 0, 1\}$: 3 orders, each with probability $(\frac{1}{4})^3$

$\{0, 1, 1\}$: 3 orders, each with probability $(\frac{1}{4})^3$

$P(X = 1) = 3 \times \frac{1}{64} + 3 \times \frac{1}{64} = \frac{6}{64} = \frac{3}{32}$ as required

3 (b) $E(X) = 0 \times \frac{1}{64} + 1 \times \frac{6}{64} + 2 \times \frac{13}{64} + 3 \times \frac{24}{64}$

$+ 4 \times \frac{13}{64} + 5 \times \frac{6}{64} + 6 \times \frac{1}{64} = \frac{192}{64} = 3$

(also clear by symmetry of the distribution about 3)

$E(X^2) = 0 \times \frac{1}{64} + 1 \times \frac{6}{64} + 4 \times \frac{13}{64} + 9 \times \frac{24}{64}$

$+ 16 \times \frac{13}{64} + 25 \times \frac{6}{64} + 36 \times \frac{1}{64}$

$E(X^2) = \frac{6 + 52 + 216 + 208 + 150 + 36}{64} = \frac{668}{64} = \frac{167}{16}$

$Var(X) = E(X^2) - [E(X)]^2 = \frac{167}{16} - 3^2 = \frac{23}{16} = 1.4375$

Question number	
3 (c)(i)	The 10 values of X must be independent.
3 (c)(ii)	$Var(Y) = 10 \times Var(X) = 10 \times \frac{23}{16} = \frac{230}{16}$
	$SD(Y) = \sqrt{\frac{230}{16}} = \frac{\sqrt{230}}{4} = 3.79$ (3 s.f.)
4 (a)	$\Sigma fx = 0 + 33 + 28 + 30 + 16 + 5 + 8 + 10 = 130$
	$\mu \approx \bar{x} = \frac{130}{100} = 1.3$
4 (b)	$C4 = P(X = 2) = e^{-1.3} \times \frac{1.3^2}{2!} = 0.2303$
	$(D4 = 100 \times 0.230289... = 23.0289)$
	$D5 = 100 - (27.2532 + 35.4291 + 23.0289) = 14.2888$
	$E3 = \frac{(33 - 35.4291)^2}{35.4291} = 0.1665$
4 (c)	H_0: the Poisson model is a good fit for the number of
	vehicles passing Eve's house per minute
	H_1: the Poisson model is not a good fit
	$X^2 = 2.8073 + 0.1665 + 3.5400 + 0.5145 = 7.0283$
	4 classes, 2 restrictions (total and estimated mean)
	$\Rightarrow \nu = 4 - 2 = 2$ degrees of freedom
	Critical value $\chi^2_2(5\%) = 5.991$
	$7.0283 > 5.991$, so reject H_0
	There is sufficient evidence at the 5% level to suggest that

Question number	
	the Poisson model is not a good fit for the number of vehicles passing Eve's house per minute.
4 (d)	4.748 < 5.991, so now H_0 is not rejected: there is insufficient evidence to suggest that the Poisson model is not a good fit. It is reasonable to remove these two values, since they are not typical: the vehicles were not travelling independently or at a constant mean rate, so they do not represent the normal situation.
5 (a)	$n = 9$, $\Sigma w = 906$, $\Sigma m = 251$, $\Sigma w^2 = 91336$, $\Sigma wm = 25388$
	$S_{ww} = 91336 - \frac{906^2}{9} = 132$, $S_{wm} = 25388 - \frac{906 \times 251}{9} = 120.666...$
	$b = \frac{120.666...}{132} = 0.91414...$
	$a = \frac{251}{9} - 0.91414... \times \frac{906}{9} = -64.134...$
	$m = 0.914w - 64.1$
5 (b)	$w = 99$: $m = 0.91414... \times 99 - 64.134... = 26.4$ g
	$w = 110$: $m = 0.91414... \times 110 - 64.134... = 36.4$ g
5 (c)	99 mm: interpolation (within 94 to 107 mm), so moderately reliable, although the points are not very close to a straight line. 110 mm: extrapolation (outside the data range) and the points are not close to a straight line, so this estimate is not reliable.
5 (d)	This is not sensible. The line $w = 0.473m + 87.5$ is the regression line of w on m, found by minimising the squares of

Question number

the w residuals, so it should only be used to estimate wing length from mass, not mass from wing length.

6 (a) The scatter diagram does not look elliptical (it is funnel shaped, with a possible outlier at (21.5, 24.8)), so the underlying distribution is probably not bivariate Normal.

6 (b)

Competitor	A	B	C	D	E	F	G	H
Rank saw	7	6	3	2	1	8	5	4
Rank chop	7	4	6	1	5	8	3	2
d	0	2	−3	1	−4	0	2	2
d^2	0	4	9	1	16	0	4	4

$\Sigma d^2 = 38$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 38}{8 \times 63} = \frac{23}{42} = 0.5476$$

6 (c) H_0: there is no association between sawing and chopping times in the population

H_1: there is positive association between sawing and chopping times in the population

$n = 8$, 5% level, one-tailed: critical value = 0.6429

$0.5476 < 0.6429$, so do not reject H_0

There is insufficient evidence at the 5% level to suggest that there is positive association between sawing and chopping times.

7 (a) X takes the values 100, 101, ..., n: that is $n - 99$ values, each with probability $\frac{1}{n-99}$

Question number

n even $\Rightarrow \frac{100+n}{2}$ is an integer, so $X < \frac{100+n}{2}$ means

$X = 100, 101, \ldots, \frac{100+n}{2} - 1$: that is $\frac{100+n}{2} - 100 = \frac{n-100}{2}$ values

$P(X < \frac{100+n}{2}) = \frac{n-100}{2(n-99)}$

7 (b) $X = W + 99$ where W is uniform on $\{1, 2, \ldots, n-99\}$

$\text{Var}(X) = \text{Var}(W) = \frac{1}{12}((n-99)^2 - 1) = \frac{1}{12}(n^2 - 198n + 9800)$

Sum of 50 independent values: variance = 50 Var(X)

$= \frac{50}{12}(n^2 - 198n + 9800) = \frac{25}{6}(n^2 - 198n + 9800)$