OCR
Oxford Cambridge and RSA

Friday 13 June 2025 – Afternoon

A Level Further Mathematics B (MEI)

Y432/01 Statistics Minor

Time allowed: 1 hour 15 minutes

187351556O

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 2025 [M/508/5596]
DC (SL) 354112/3

OCR is an exempt Charity

stemup.co.uk

1 A discrete random variable, X, has a Poisson distribution with mean 4.

(a) Write down the value of $\mathrm{Var}(X)$. **[1]**

(b) Find $\mathrm{P}(X = 3)$. **[1]**

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

Another discrete random variable, Y, which is independent of X, has a Poisson distribution with mean 5.

(d) Find $\mathrm{P}(X + Y \leqslant 9)$. **[2]**

2 In an experiment, the speed of an object, $v\,\mathrm{m\,s^{-1}}$, at time t seconds, is recorded at intervals of 10 seconds from $t = 0$ to $t = 80$.

The summary statistics are as follows.

$n = 8$, $\Sigma t = 280$, $\Sigma t^2 = 14000$, $\Sigma v = 876.28$, $\Sigma v^2 = 133310.2$, $\Sigma tv = 43190$

A linear function of t is used as a model for v.

(a) By first determining the equation of an appropriate regression line, estimate the value of v when $t = 45$. **[3]**

(b) Comment on the reliability of your estimate in part **(a)**. **[1]**

(c) Explain why in this scenario it would be inappropriate to use a value of v to estimate a value of t using **any** regression line. **[1]**

© OCR 2025

stemup.co.uk

3 An archer fires arrows one by one at a target until an arrow hits the target. The number of arrows fired up to and including the arrow that hits the target is denoted by X. The archer considers using a geometric model for the distribution of X.

(a) State **two** assumptions required for such a model to be appropriate. **[2]**

For the remainder of this question you may assume that a geometric model is appropriate for the distribution of X.

The archer keeps records of the values of X. He collects a sample of these values and uses them to estimate the variance of X.

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

The estimate of the variance of X calculated from the sample is 12.

(c) Determine an estimate of the probability that when the archer fires a single arrow at the target, the arrow hits the target. **[3]**

(d) Use your answer to part **(c)** to write down an estimate of the expected number of arrows that the archer must fire at the target up to and including the first arrow that hits the target. **[1]**

4 The discrete random variable, X, has a uniform distribution over the values $\{5, 6, ..., n\}$.

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

In part **(a)** you should assume that $n = 24$.

(i) Using the formula $\mathrm{E}(X) = \sum x_i p_i$, determine the value of $\mathrm{E}(X)$ by direct calculation. **[3]**

(ii) Use the substitution $Y = X - 4$ and the given formula in the formulae booklet to determine $\mathrm{Var}(X)$. **[2]**

(iii) Show that $\mathrm{P}(X > \mathrm{E}(X)) = \frac{1}{2}$. **[1]**

(b) A student states that $\mathrm{P}(X > \mathrm{E}(X)) = \frac{1}{2}$ for **all** $n \geqslant 24$.

Explain whether the student is correct. **[1]**

© OCR 2025

stemup.co.uk

5 Long-term data collected at a certain farm indicates that 30% of eggs that it produces are large and the rest are categorised as medium.

Eggs are collected in batches of 4 and the number of large eggs in each batch, X, is recorded.

It is suggested that a binomial distribution might be appropriate to model X.

(a) (i) State **two** reasons, within the context of this question, why a binomial model might be appropriate to model X. **[2]**

(ii) State **two** assumptions, within the context of this question, that must be made for a binomial model to be appropriate to model X. You should not refer to the reasons given in part **(a)(i)**. **[2]**

(iii) Write down the binomial distribution which might be a suitable model for the distribution of X. **[1]**

The farm manager investigates whether the distribution in part **(a)(iii)** is a good model for X.

(b) State **one** reason why taking data from a sample of 6 batches would **not** be appropriate to carry out the investigation. **[1]**

The farm manager takes a random sample of 100 batches. The results are summarised in **Table 5.1**.

Table 5.1

x	0	1	2	$\geqslant 3$
Frequency	31	36	30	3

Table 5.2, shows expected frequencies for the values of X in 100 batches, assuming that X has the distribution from part **(a)(iii)**. The table is incomplete.

Table 5.2

x	0	1	2	$\geqslant 3$
Expected frequency	24.01			

(c) Complete the copy of **Table 5.2** in the **Printed Answer Booklet**. Give the expected frequencies to **2** decimal places. **[2]**

(d) Carry out a χ^2 test, at the 5% significance level, for goodness of fit of the model in part **(a)(iii)**. **[5]**

© OCR 2025

stemup.co.uk

6 A factory discharges very low concentrations of an inert chemical through an outfall into a nearby body of water.

The concentration of the chemical is measured by sensors at seven known locations at fixed distances from the outfall. Each measurement is known to be affected by various environmental factors. None of these factors are measured or noted.

The manager of the factory needs to assess the relationship between the concentration of the chemical at each sensor and the distance of the sensor from the outfall.

The manager can carry out a hypothesis test **either** for any correlation **or** for any association between the concentration of the chemical at the sensor and the distance of the sensor from the outfall.

(a) Explain why the manager should carry out a test for association rather than one for correlation. **[1]**

The concentration of the chemical at each sensor and the distance of each sensor from the outfall, both in arbitrary units, for a particular day after the chemical has been discharged are given in the table below.

Sensor	A	B	C	D	E	F	G
Distance from outfall (arbitrary units)	1.2	6.6	8.9	11.0	15.6	23.2	30.5
Concentration of chemical (arb. units)	621	459	498	610	222	453	420

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

Carry out the appropriate test for association at the 5% significance level. A copy of the table, with spare, empty rows, is given in the **Printed Answer Booklet** to help with any calculations. **[8]**

Turn over for question 7

stemup.co.uk

7 The probability function of the discrete random variable, X, is shown in the table. You are given that α, β and γ are constants.

x	-1	0	1
$\mathrm{P}(X = x)$	α	β	γ

(a) Find $\mathrm{E}(X)$ in terms of α and γ. **[1]**

(b) Find $\mathrm{Var}(X)$ in terms of α and γ. **[2]**

The random variable W is defined by $W = X_1 + X_2$ where X_1 and X_2 are independent random variables each representing an observation of X.

(c) Determine the probability distribution of W in terms of α, β and γ. Your answer should be entered in the empty table in the **Printed Answer Booklet**. You should not need all of the columns. **[4]**

(d) By forming an expression for $\mathrm{E}(W)$ from the table found in **(c)** and then eliminating β, show that $\mathrm{E}(W) = 2\mathrm{E}(X)$. **[4]**

(e) In the case where $\alpha = 0.1$ and $\beta = 0.7$, find $\mathrm{P}(W = 0 \,|\, X_1 = 1)$. **[1]**

END OF QUESTION PAPER

© OCR 2025

stemup.co.uk

BLANK PAGE

© OCR 2025

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 2025

stemup.co.uk

OCR
Oxford Cambridge and RSA

Friday 13 June 2025 – Afternoon

A Level Further Mathematics B (MEI)

Y432/01 Statistics Minor

Printed Answer Booklet

Time allowed: 1 hour 15 minutes

AB

You must have:
- Question Paper Y432/01 (inside this document)
- the Formulae Booklet for Further Mathematics B (MEI)
- a scientific or graphical calculator

Please write clearly in black ink. **Do not write in the barcodes.**

Centre number | | | | | | Candidate number | | | | |

First name(s) StemUp Model Solution

Last name

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.
- 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.

INFORMATION
- This document has **12** pages.

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

© OCR 2025 [M/508/5596]
DC (SL) 354115/1
OCR is an exempt Charity

stemup.co.uk

1(a)	$\text{Var}(X) = 4$ (mean = variance for a Poisson distribution)
1(b)	$P(X = 3) = e^{-4} \times 4^3/3!$
	$= 0.195$ (3 s.f.)
1(c)	$P(X > 3) = 1 - P(X \leqslant 3)$
	$= 1 - 0.43347...$
	$= 0.567$ (3 s.f.)
1(d)	X and Y independent Poisson, so $X + Y \sim \text{Po}(4 + 5) = \text{Po}(9)$
	$P(X + Y \leqslant 9) = 0.58741...$
	$= 0.587$ (3 s.f.)
2(a)	t is controlled, so use the regression line of v on t.
	$S_{tt} = \Sigma t^2 - (\Sigma t)^2/n = 14000 - 280^2/8 = 4200$
	$S_{tv} = \Sigma tv - \Sigma t \Sigma v/n = 43190 - (280 \times 876.28)/8 = 12520.2$
	$b = S_{tv} \div S_{tt} = 12520.2 \div 4200 = 2.981$
	$\bar{t} = 280/8 = 35, \quad \bar{v} = 876.28/8 = 109.535$
	$v - 109.535 = 2.981(t - 35)$
	$v = 2.981t + 5.2$
	When $t = 45$: $v = 2.981 \times 45 + 5.2 = 139.345$
	$v \approx 139$ m s^{-1}
2(b)	$t = 45$ lies within the range of the data ($t = 0$ to 80), so this is
	interpolation and the estimate is likely to be reliable.

© OCR 2025

2(c)

t is not a random variable: it is controlled (the times are fixed in advance at intervals of 10 s). The regression line of t on v is only valid if t is random, so no regression line should be used to estimate t from a value of v.

3(a)

1. Whether one arrow hits the target is independent of whether any other arrow hits the target.
2. The probability that an arrow hits the target is the same (constant) for every arrow fired.

3(b)

1. The sample should be large.
2. The sample should be random (representative of all of the archer's values of X).

3(c)

Let p = P(an arrow hits the target), $X \sim Geo(p)$

$Var(X) = (1 - p)/p^2 = 12$

$1 - p = 12p^2$

$12p^2 + p - 1 = 0$

$(4p - 1)(3p + 1) = 0$

$p = 1/4$ or $p = -1/3$ (reject, since $p > 0$)

Estimate of probability = 0.25

3(d)

$E(X) = 1/p = 1/0.25 = 4$

Expected number of arrows = 4

© OCR 2025

stemup.co.uk

4(a)(i)	X takes the 20 values 5, 6, ..., 24, each equally likely
	$P(X = x) = 1/20$ for $x = 5, 6, ..., 24$
	$E(X) = \Sigma\, x\, p_i = 5 \times 1/20 + 6 \times 1/20 + 7 \times 1/20 + ... + 24 \times 1/20$
	$= (1/20) \times (5 + 6 + 7 + ... + 24)$
	$5 + 6 + ... + 24$ is an arithmetic series with 20 terms:
	$= (20/2) \times (5 + 24) = 10 \times 29 = 290$
	$E(X) = 290/20$
	$E(X) = 14.5$
4(a)(ii)	$Y = X - 4$ is uniform on $\{1, 2, ..., 20\}$, so $n = 20$ in the formula
	$Var(Y) = (n^2 - 1)/12 = (20^2 - 1)/12 = 399/12 = 33.25$
	$Var(X) = Var(Y + 4) = Var(Y)$, since adding a constant
	does not change the variance
	$Var(X) = 33.25$
4(a)(iii)	$P(X > E(X)) = P(X > 14.5) = P(X \geqslant 15)$
	$X \in \{15, 16, ..., 24\}$: that is $24 - 15 + 1 = 10$ values
	$P(X > E(X)) = 10/20 = 1/2$ as required
4(b)	No. e.g. $n = 25$: $E(X) = (5 + 25)/2 = 15$ and there are 21 values,
	$X > 15$ means $X \in \{16, ..., 25\}$, 10 values, so $P(X > E(X)) = 10/21 \neq 1/2$
	(It only holds when n is even, so the student is not correct.)

© OCR 2025

BLANK PAGE

DO NOT WRITE ON THIS PAGE

© OCR 2025

stemup.co.uk

5(a)(i)

1. There is a fixed number of eggs (4) in each batch.
2. Each egg is either large or medium (two outcomes), and
   we are counting the number of large eggs in the batch.

5(a)(ii)

1. Whether one egg is large or medium is independent of
   whether any other egg in the batch is large or medium.
2. The probability that an egg is large is the same (0.3)
   for every egg.

5(a)(iii)

$X \sim B(4, 0.3)$

5(b)

6 batches is too small a sample to be representative of the population, and the expected frequencies would be too small (less than 5) to carry out a χ^2 goodness of fit test.

5(c)

Expected frequency = $100 \times P(X = x)$, $X \sim B(4, 0.3)$

$x = 1$: $100 \times 4 \times 0.3 \times 0.7^3 = 41.16$

$x = 2$: $100 \times 6 \times 0.3^2 \times 0.7^2 = 26.46$

$x \geqslant 3$: $100 - 24.01 - 41.16 - 26.46 = 8.37$

x	0	1	2	$\geqslant 3$
Expected frequency	24.01	41.16	26.46	8.37

© OCR 2025

5(d)

H_0: B(4, 0.3) is a good model for the number of large eggs in a batch

H_1: B(4, 0.3) is not a good model for the number of large eggs

x	O	E	$(O - E)^2/E$
0	31	24.01	2.0350
1	36	41.16	0.6469
2	30	26.46	0.4736
≥ 3	3	8.37	3.4453

All expected frequencies are at least 5, so no cells need combining.

Test statistic $X^2 = \Sigma(O - E)^2/E = 2.0350 + 0.6469 + 0.4736 + 3.4453$

$X^2 = 6.601$ (4 s.f.)

Degrees of freedom = 4 − 1 = 3 (no parameters estimated from the data)

Critical value at 5% level: $\chi^2_3(5\%) = 7.815$

6.601 < 7.815, so the test statistic is not in the critical region.

Do not reject H_0.

There is insufficient evidence at the 5% level to suggest that

B(4, 0.3) is not a good model for the number of large eggs in a batch.

© OCR 2025

stemup.co.uk

6(a)

The distances are fixed (not random), so the data cannot come from a bivariate Normal distribution, and the relationship need not be linear.

6(b)

Sensor	A	B	C	D	E	F	G
Distance from outfall (arbitrary units)	1.2	6.6	8.9	11.0	15.6	23.2	30.5
Concentration of chemical (arb. units)	621	459	498	610	222	453	420
Rank of distance	1	2	3	4	5	6	7
Rank of concentration	7	4	5	6	1	3	2
d^2	36	4	4	4	16	9	25

(ranks of concentration from smallest = 1; d = difference in ranks)

$\Sigma d^2 = 36 + 4 + 4 + 4 + 16 + 9 + 25 = 98$

$r_s = 1 - 6\Sigma d^2 \div [n(n^2 - 1)]$

$= 1 - (6 \times 98)/(7 \times (49 - 1)) = 1 - 588/336$

$r_s = -0.75$

H_0: there is no association between the distance of a sensor from the outfall and the concentration of chemical at the sensor in the population

H_1: there is association between the distance of a sensor from the outfall and the concentration of chemical at the sensor in the population

Two-tailed test, n = 7, 5% significance level:

critical value = 0.7857

$|r_s| = 0.75 < 0.7857$ $(-0.7857 < -0.75 < 0.7857)$

so r_s is not in the critical region.

Do not reject H_0.

There is insufficient evidence at the 5% level to suggest that there is association between the distance of a sensor from the outfall and the concentration of the chemical at the sensor.

(answer space continued on next page)

© OCR 2025

6(b)	(continued)

© OCR 2025

stemup.co.uk

7(a)

$E(X) = (-1)a + 0 \times b + 1 \times c$

$E(X) = c - a$

7(b)

$E(X^2) = (-1)^2a + 0^2b + 1^2c = a + c$

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

$\text{Var}(X) = a + c - (c - a)^2$

$(= a + c - a^2 + 2ac - c^2)$

7(c)

$W = -2$: both -1, prob a^2; $W = 2$: both 1, prob c^2

$W = -1$: $(-1, 0)$ or $(0, -1)$, prob $2ab$; $W = 1$: $(0, 1)$ or $(1, 0)$, prob $2bc$

$W = 0$: $(0, 0)$, $(-1, 1)$ or $(1, -1)$, prob $b^2 + 2ac$

w	-2	-1	0	1	2		
$P(W = w)$	a^2	$2ab$	$b^2 + 2ac$	$2bc$	c^2		

7(d)

$E(W) = (-2)a^2 + (-1)(2ab) + 0(b^2 + 2ac) + 1(2bc) + 2c^2$

$= 2c^2 - 2a^2 + 2bc - 2ab$

$= 2(c - a)(c + a) + 2b(c - a)$

Probabilities sum to 1: $a + b + c = 1$, so $b = 1 - a - c$

$E(W) = 2(c - a)(c + a) + 2(1 - a - c)(c - a)$

$= 2(c - a)[(c + a) + (1 - a - c)]$

$= 2(c - a) \times 1$

$= 2(c - a)$

$= 2E(X)$ as required

7(e)

$c = 1 - 0.1 - 0.7 = 0.2$

$W = 0$ with $X_1 = 1$ needs $X_2 = -1$

$P(W = 0 \mid X_1 = 1) = P(X_1 = 1 \text{ and } X_2 = -1) \div P(X_1 = 1)$

$= (0.2 \times 0.1) \div 0.2 = 0.1$

© OCR 2025

EXTRA ANSWER SPACE

If you need extra space use these lined pages. You must write the question numbers clearly in the margin.

© OCR 2025

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 2025

stemup.co.uk