OCR
RECOGNISING ACHIEVEMENT

Friday 25 January 2013 – Afternoon

AS GCE MATHEMATICS (MEI)

4766/01 Statistics 1

QUESTION PAPER

4732980113

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4766/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

This paper has been pre modified for carrier language

DC (LEG/CGW) 61801/3

Section A (36 marks)

1 The stem and leaf diagram illustrates the heights in metres of 25 young oak trees.

3	4 6 7 8 9 9
4	0 2 2 3 4 6 8 9
5	0 1 3 5 8
6	2 4 5
7	4 6
8	1

Key: 4 | 2 represents 4.2

(i) State the type of skewness of the distribution. **[1]**

(ii) Use your calculator to find the mean and standard deviation of these data. **[3]**

(iii) Determine whether there are any outliers. **[4]**

2 The probability distribution of the random variable X is given by the formula

$$\mathrm{P}(X = r) = k(r^2 - 1) \text{ for } r = 2, 3, 4, 5.$$

(i) Show the probability distribution in a table, and find the value of k. **[3]**

(ii) Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. **[5]**

3 Each weekday Alan drives to work. On his journey, he goes over a level crossing. Sometimes he has to wait at the level crossing for a train to pass.

- W is the event that Alan has to wait at the level crossing.
- L is the event that Alan is late for work.

You are given that $\mathrm{P}(L \mid W) = 0.4$, $\mathrm{P}(W) = 0.07$ and $\mathrm{P}(L \cup W) = 0.08$.

(i) Calculate $\mathrm{P}(L \cap W)$. **[2]**

(ii) Draw a Venn diagram, showing the events L and W. Fill in the probability corresponding to each of the four regions of your diagram. **[3]**

(iii) Determine whether the events L and W are independent, explaining your method clearly. **[3]**

4 At a dog show, three out of eleven dogs are to be selected for a national competition.

(i) Find the number of possible selections. **[2]**

(ii) Five of the eleven dogs are terriers. Assuming that the dogs are selected at random, find the probability that at least two of the three dogs selected for the national competition are terriers. **[5]**

5 Malik is playing a game in which he has to throw a 6 on a fair six-sided die to start the game. Find the probability that

(i) Malik throws a 6 for the first time on his third attempt, **[3]**

(ii) Malik needs at most ten attempts to throw a 6. **[2]**

Section B (36 marks)

6 The heights x cm of 100 boys in Year 7 at a school are summarised in the table below.

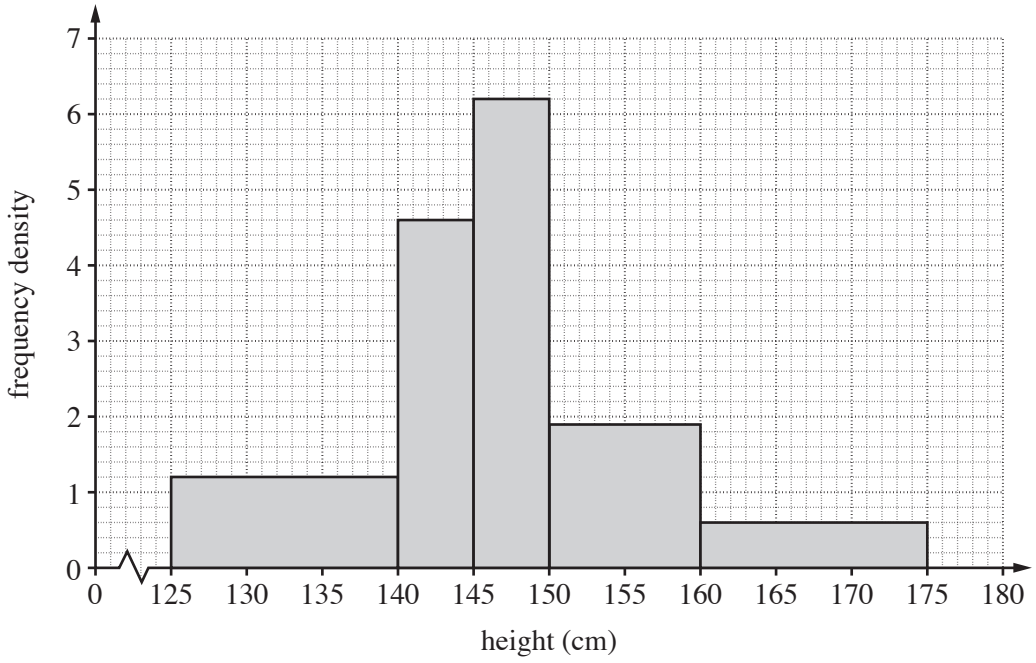
Height	$125 \leqslant x \leqslant 140$	$140 < x \leqslant 145$	$145 < x \leqslant 150$	$150 < x \leqslant 160$	$160 < x \leqslant 170$
Frequency	25	29	24	18	4

(i) Estimate the number of boys who have heights of at least 155 cm. **[2]**

(ii) Calculate an estimate of the median height of the 100 boys. **[3]**

(iii) Draw a histogram to illustrate the data. **[5]**

The histogram below shows the heights of 100 girls in Year 7 at the same school.



(iv) How many more girls than boys had heights exceeding 160 cm? **[3]**

(v) Calculate an estimate of the mean height of the 100 girls. **[5]**

7 A coffee shop provides free internet access for its customers. It is known that the probability that a randomly selected customer is accessing the internet is 0.35, independently of all other customers.

(i) 10 customers are selected at random.

(A) Find the probability that exactly 5 of them are accessing the internet. **[3]**

(B) Find the probability that at least 5 of them are accessing the internet. **[2]**

(C) Find the expected number of these customers who are accessing the internet. **[2]**

Another coffee shop also provides free internet access. It is suspected that the probability that a randomly selected customer at this coffee shop is accessing the internet may be different from 0.35. A random sample of 20 customers at this coffee shop is selected. Of these, 10 are accessing the internet.

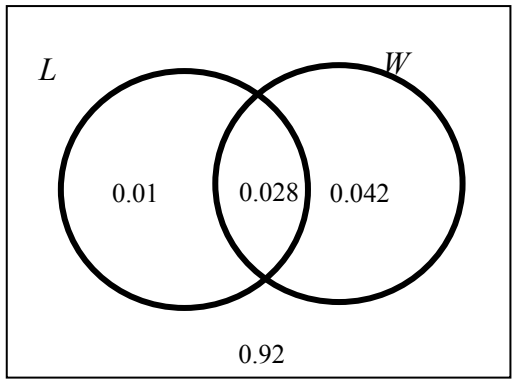
(ii) Carry out a hypothesis test at the 5% significance level to investigate whether the probability for this coffee shop is different from 0.35. Give a reason for your choice of alternative hypothesis. **[9]**

(iii) To get a more reliable result, a much larger random sample of 200 customers is selected over a period of time, and another hypothesis test is carried out. You are given that 90 of the 200 customers were accessing the internet. You are also given that, if X has the binomial distribution with parameters $n = 200$ and $p = 0.35$, then $\mathrm{P}(X \geqslant 90) = 0.0022$. Using the same hypotheses and significance level which you used in part **(ii)**, complete this test. **[2]**

6 (iii)

Question			Answer	Marks	Guidance	
1	(i)		Positive	B1 **[1]**	CAO	
1	(ii)		Mean = 5.064 allow 5.1 with working 126.6/25 or 5.06 without SD = 1.324 allow 1.3 with working or 1.32 without	B1 B2 **[3]**	Allow B1 for RMSD = 1.297 or var =1.753 or MSD = 1.683	Also allow B1 for Sxx = 42.08 or for $\Sigma x^2 = 683$ SC1 for both mean = 50.64 and SD = 13.24 (even if over-specified)
1	(iii)		$\bar{x} - 2s = 5.064 - 2\times1.324 = 2.416$ $\bar{x} + 2s = 5.064 + 2\times1.324 = 7.712$ So there is an outlier.	B1FT M1 A1FT E1 **[4]**	FT their mean and sd for $\bar{x} + 2s$ but withhold final E mark if their limits mean that there are no outliers. For upper limit Incorrect statement such as 7.6 and 8.1 are outliers gets E0 Do not award E1 if calculation error in upper limit	For use of quartiles and IQR $Q_1 = 3.95$; $Q_3 = 6.0$; IQR = 2.05 3.95 – 1.5(2.05) gets M1 Allow other sensible definitions of quartiles 6.0 + 1.5(2.05) gets M1 Limits 0.875 and 9.075 So there are no outliers NB do not penalise over-specification here as not the final answer but just used for comparison. FT from SC1
2	(i)		<table><tr><td>r</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>P($X = r$)</td><td>3k</td><td>8k</td><td>15k</td><td>24k</td></tr></table> $3k + 8k + 15k + 24k = 1$ $k = 0.02$	B1 M1 A1 **[3]**	For correct table (ito k or correct probabilities 0.06, 0.16, 0.30, 0.48) or $k = 1/50$ (with or without working)	 For their four multiples of k added and =1. Allow M1A1 even if done in part **(ii)** – link part **(ii)** to part **(i)**

Question			Answer	Marks	Guidance	
2	(ii)		$E(X) = (2\times0.06)+(3\times0.16)+(4\times0.30)+(5\times0.48) = 4.2$	M1	For Σrp (at least 3 terms correct Provided 4 reasonable probabilities seen.	If probs wrong but sum = 1 allow full marks here. If sum ≠ 1 allow max M1A0M1 M0A0 (provided all probabilities between 0 and 1)
			or 21/5	A1	cao	Or ito k NB $E(X) = 210k$, $E(X^2) = 924k$ gets M1A0M1M0A0. $E(X) = 210k$, Var $(X) = 924k – (210k)^2$ gets M1A0M1M1A0.
			$E(X^2) = (4\times0.06)+(9\times0.16)+(16\times0.30)+(25\times0.48) = 18.48$	M1	For Σr^2p (at least 3 terms correct)	
			$Var(X) = 18.48 – 4.2^2$	M1	dep for – their $E(X)^2$	
			$= 0.84 = 21/25$	A1	FT their $E(X)$ provided $Var(X) > 0$ (and of course $E(X^2)$ is correct)	Use of $E(X-\mu)^2$ gets M1 for attempt at $(x-\mu)^2$ should see $(–2.2)^2$, $(–1.2)^2$, $(–0.2)^2$, 0.8^2, (if $E(X)$ wrong FT their $E(X)$) (all 4 correct for M1), then M1 for $\Sigma p(x-\mu)^2$ (at least 3 terms correct with their probabilities) Division by 4 or other spurious value at end gives max M1A1M1M1A0, or M1A0M1M1A0 if $E(X)$ also divided by 4. Unsupported correct answers get 5 marks
				[5]		
3	(i)		$P(L\cap W) = P(L\mid W)\times P(W) = 0.4\times0.07 = 0.028$	M1	For $P(L\mid W)\times P(W)$	
				A1	cao	
				[2]		

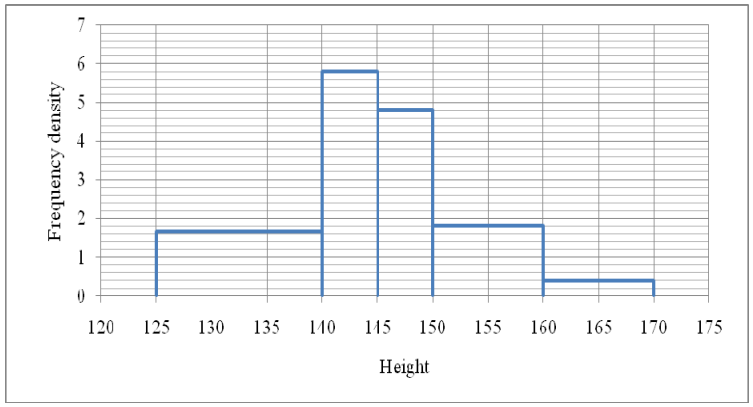
Question			Answer	Marks	Guidance	
3	(ii)			B1 B1 B1 **[3]**	For two labelled intersecting circles For at least 2 correct probabilities. For remaining probabilities	FT their 0.028 provided < 0.038
3	(iii)		$P(L \cap W) = 0.028, P(L) \times P(W) = 0.038 \times 0.07 = 0.00266$ Not equal so not independent	M1 A1 E1* dep on M1 **[3]**	For correct use of $P(L) \times P(W)$ If P(L) wrong, max M1A0E0. No marks if P(W) wrong For 0.00266 Allow 'they are dependent' Do not award E1 if P($L \cap W$)wrong	Or EG $P(L \mid W) = 0.4, P(L) = 0.038$ Not equal so not independent M1 is for comparing with some attempt at numbers P($L \mid W$) with P(L) , A1 for 0.038 If P(L) wrong, max M1A0E0
4	(i)		$\binom{11}{3}$ = 165	M1 A1 **[2]**	Seen Cao	

Question			Answer	Marks	Guidance	
4	(ii)		$\frac{\binom{5}{2}\times\binom{6}{1}}{\binom{11}{3}}+\frac{\binom{5}{3}\times\binom{6}{0}}{\binom{11}{3}}=\frac{60}{165}+\frac{10}{165}=\frac{70}{165}=\frac{14}{33}=0.424$	M1	For intention to add correct two fractional terms	**Or** For attempt at correct two terms
				M1	For numerator of first term	For prod of 3 correct fractions =4/33
			Alternative	M1	For numerator of sec term	For whole expression ie
			$1-\text{P(1 or 0)}=1-3\times\frac{5}{11}\times\frac{6}{10}\times\frac{5}{9}-\frac{6}{11}\times\frac{5}{10}\times\frac{4}{9}$		Do not penalise omission of $\binom{6}{0}$	$3\times\frac{5}{11}\times\frac{4}{10}\times\frac{6}{9}\left(=\frac{4}{11}\right)$ (= 3 × 0.1212…)
			$=1-\frac{5}{11}-\frac{4}{33}=\frac{14}{33}$			
			M1 for 1 – P(1 or 0) , M1 for first product, M1 for ×3, M1 for second product, A1			
				M1	For correct denominator	For attempt at $\frac{5}{11}\times\frac{4}{10}\times\frac{3}{9}\left(=\frac{2}{33}\right)$
				A1	cao	cao
				[5]		Use of binomial can get max first M1
5	(i)		$\left(\frac{5}{6}\right)^2\times\frac{1}{6}=\frac{25}{216}$ (= 0.116)	M1	For 5/6 (or 1 – 1/6) seen	If extra term or whole number factor present give M1M0A0
				M1	For whole product	
				A1	cao	Allow 0.12 with working
				[3]		
5	(ii)		$1-\left(\frac{5}{6}\right)^{10}=1-0.1615=0.8385$	M1	For $(5/6)^{10}$ (without extra terms)	Allow 0.838 or 0.839 without working and 0.84 with working. For addition P(X = 1) +…+ P(X = 10) give M1A1 for 0.84 or better, otherwise M0A0
				A1	cao	
				[2]		

Question			Answer	Marks	Guidance	
6	(i)		4 + ½ of 18 = 4 + 9 =13	M1 A1 **[2]**	For ½ of 18 cao	13/100 gets M1A0
6	(ii)		(Median) = 50.5th value	M1	For 50.5 seen	SC2 for use of 50th value leading to Est = 140 + (25 /29 × 5) = 144.3 (SC1 if over-specified)
			Est = $140 + \left(\frac{25.5}{29}\right)\times 5$ or = $140 + \left(\frac{50.5-25}{54-25}\right)\times 5$	M1	For attempt to find this value	or Est = $145 - \left(\frac{3.5}{29}\right)\times 5$ = 144.4
			= 144.4	A1 **[3]**		NB no marks for mean = 144.35 NB Watch for over-specification

Question	Answer	Marks	Guidance
6 (iii)			

Height	Frequency	Group width	Frequency density
$125 \leq x \leq 140$	25	15	1.67
$140 < x \leq 145$	29	5	5.80
$145 < x \leq 150$	24	5	4.80
$150 < x \leq 160$	18	10	1.80
$160 < x \leq 170$	4	10	0.40



Marks		Guidance
M1	For fd's - at least 3 correct Accept any suitable unit for fd such as eg freq per cm.	M1 can be also be gained from freq per 10 – 16.7, 58, 48, 18, 4 (at least 3 correct) or freq per 5 – 8.35, 29, 24, 9, 2
A1	correct to at least one dp allow 1.66 but not 1.6 for first fd	for all correct. If fd not explicitly given, M1 A1 can be gained from all heights correct (within one square) on histogram (and M1A0 if at least 3 correct)
G1	linear scales on both axes and label on vertical axis	Linear scale and label on vertical axis IN RELATION to first M1 mark ie fd or frequency density or if relevant freq/10, etc (NOT eg fd/10). However allow scale given as fd×10, or similar Accept f/w or f/cw (freq/width or freq/class width) Can also be gained from an accurate key G0 if correct label but not fd's.
W1	width of bars	Must be drawn at 125, 140 etc NOT 124.5 or 125.5 etc NO GAPS ALLOWED Must have linear scale. No inequality labels on their own such as 125≤S<140, etc but allow if a clear horizontal linear scale is also given. Ignore horizontal label.
H1	height of bars	Height of bars – must be linear vertical scale. FT of heights dep on at least 3 heights correct and all must agree with their

<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th>Guidance</th><th></th></tr>
<tr><td></td><td></td><td></td><td></td><td>[5]</td><td></td><td>fds
If fds not given and at least 3 heights correct then max M1A0G1W1H0
Allow restart with correct heights if given fd wrong (for last three marks only)</td></tr>
<tr><td>6</td><td>(iv)</td><td></td><td>4 boys
0.6 × 15

= 9 girls
So 5 more girls</td><td>M1

A1
A1
[3]</td><td>For 0.6 × 15

For 9 girls
cao</td><td>Or 45 × 0.2 = 9 (number of squares and 0.2 per square)</td></tr>
<tr><td>6</td><td>(v)</td><td></td><td>Frequencies and midpoints for girls are
<table>
<tr><td>Height</td><td>132.5</td><td>142.5</td><td>147.5</td><td>155</td><td>167.5</td></tr>
<tr><td>Frequency</td><td>18</td><td>23</td><td>31</td><td>19</td><td>9</td></tr>
</table>
So mean = $\frac{(132.5\times18)+(142.5\times23)+(147.5\times31)+(155\times19)+(167.5\times9)}{100}$
$=\frac{(2385)+(3277.5)+(4572.5)+(2945)+(1507.5)}{100}$
= 146.9
(Exact answer 146.875)</td><td>B1

B1

M1
M1*
Dep on M1

A1

[5]</td><td>For at least three frequencies correct

At least three midpoints correct

For attempt at $\sum xf$
For division by 100

Cao

NB Watch for over-specification</td><td>No further marks if not using midpoints

For sight of at least 3 xf pairs

Allow answer 146.9 or 147 but not 150
NB Accept answers seen without working (from calculator)
Use of ‘not quite right’ midpoints such as 132.49 or 132.51 etc can get B1B0M1M1A0</td></tr>
</table>

Question			Answer	Marks	Guidance	
7	(i)	(*A*)	$X \sim$ B(10, 0.35) P(5 accessing internet) = $\binom{10}{5} \times 0.35^5 \times 0.65^5$ = 0.1536 **OR** from tables = 0.9051 – 0.7515 = 0.1536	M1 M1 A1 **OR** M2 A1 **[3]**	or $0.35^5 \times 0.65^5$ For $\binom{10}{5} \times p^5 \times q^5$ cao For 0.9051 – 0.7515 cao	With p + $\boldsymbol{q}$ = **1** Also for 252 × 0.0006094 Allow 0.15 or better NB 0.153 gets A0 See tables at the website http://www.mei.org.uk/files/pdf/formula_book_mf2.pdf
7	(i)	(*B*)	P($X \geq 5$) = 1 – P($X \leq 4$) =1 – 0.7515 = 0.2485	 M1 A1 **[2]**	 For 0.7515 cao	 Accept 0.25 or better – allow 0.248 or 0.249 Calculation of individual probabilities gets B2 if fully correct 0.25 or better, otherwise B0.
7	(i)	(*C*)	E(X) = np = 10 × 0.35 = 3.5	M1 A1 **[2]**	For 10 × 0.35 cao	If any indication of rounding to 3 or 4 allow M1A0

Question			Answer	Marks	Guidance	
7	(ii)		Let $X \sim$ B(20, 0.35) Let p = probability of a customer using the internet (for population)	B1	For definition of p in context	Minimum needed for B1 is p = probability of using internet. Allow p = P(using internet) Definition of p must include word probability (or chance or proportion or percentage or likelihood but NOT possibility). Preferably as a separate comment. However can be at end of H_0 as long as it is a clear definition 'p = the probability of using internet', Do NOT allow 'p = the probability of using internet is different'
			H_0: $p = 0.35$	B1	For H_0	Allow p=35%, allow only p or θ or π or ρ. However allow any single symbol <u>if defined</u> (including x) Allow $H_0 = p$=0.35, Allow H_0 : $p={}^{7}/_{20}$or $p={}^{35}/_{100}$ Allow NH and AH in place of H_0 and H_1 Do not allow H_0 : P(X=x) = 0.35 Do not allow H_0: =0.35, =35%, P(0.35), p(x)=0.35, x=0.35 (unless x correctly defined as a probability) Do not allow H_0 and H_1 reversed For hypotheses given in words allow Maximum B0B1B1 Hypotheses in words must include probability (or chance or proportion or percentage) and the figure 0.35 oe Thus eg H_0 : p(using internet) = 0.35, H_1 : p(using internet) $\neq$ 0.35 gets B0B1B1

Question			Answer	Marks	Guidance	
			H_1: $p \neq 0.35$	B1	For H_1	Allow '$p < 0.35$ or $p > 0.35$' in place of $p \neq 0.35$
			H_1 has this form because the test is to investigate whether the proportion is different, (rather than lower or higher).	E1		Do not allow if H_1 wrong.
			$P(X \geq 10)$	B1	For notation $P(X \geq 10)$ or $P(X > 9)$ or $1 - P(X \leq 9)$ (as long as no incorrect notation)	This mark may be implied by 0.1218 as long as no incorrect notation. No further marks if point probs used - $P(X = 10) = 0.0686$ (do not even give the notation mark for correct notation) DO NOT FT wrong H_1, but see extra notes
			$= 1 - 0.8782 = 0.1218$	B1*	For 0.1218 Allow 0.12	Or for $1 - 0.8782$ Indep of previous mark
			$> 2.5\%$	M1* dep	For comparison with 2.5%	
			So not significant.	A1*		Allow 'accept H_0' or 'reject H_1'
			Conclude that there is not enough evidence to indicate that the probability is different. (Must state 'probability', not just 'p')	E1* dep on A1		Must include 'sufficient evidence' or something similar such as 'to suggest that' ie an element of doubt either in the A or E mark.
			ALTERNATIVE METHOD FOR FINAL 5 MARKS			
			Critical region method			Do not insist on correct notation as candidates have to work out two probabilities for full marks. If only upper tail of CR given (or only upper tail justified), allow max 4/5 for final 5 marks.
			LOWER TAIL $P(X \leq 2) = 0.0121 < 2.5\%$ $P(X \leq 3) = 0.0444 > 2.5\%$	B1	For either probability	
			UPPER TAIL $P(X \geq 11) = 1 - P(X \leq 10) = 1 - 0.9468 = 0.0532 > 2.5\%$ $P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.9804 = 0.0196 < 2.5\%$	B1	For either probability	

Question			Answer	Marks	Guidance	
			So critical region is {0,1,2,12,13,14,15,16,17,18,19,20}	M1* dep	cao dep on at least one correct comparison with 2.5%	No marks if CR not justified Condone {0,1,2, 12, … 20}, $X \leq 2$, $X \geq 12$, oe but not P($X \leq 2$) etc
			So not significant Conclude that there is not enough evidence to indicate that the probability is different.	A1* E1* dep on A1 **[9]**		NB If CR found correctly then P($X = 10$) subsequently found but cand says '10 not in CR' then allow up to all last five marks. If do not say '10 not in CR' allow none of last five marks
7	**(iii)**		0.0022 < 2.5% So reject H_o, Significant.	B1	For either reject H_o or significant, dep on correct comparison	
			Conclude that there is enough evidence to indicate that the probability is different.	E1* dep **[2]**	Dep on good attempt at correct hypotheses in part (ii)	If they have H_1: p>0.35, allow SC1 if all correct including comparison with 5%.