OCR
RECOGNISING ACHIEVEMENT

Thursday 6 June 2013 – Morning

A2 GCE MATHEMATICS

4734/01 Probability & Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4734/01
- List of Formulae (MF1)

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.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Give non-exact numerical answers correct to 3 significant figures unless a different degree of accuracy is specified in the question or is clearly appropriate.

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 reminded of the need for clear presentation in your answers.**
- 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.

1 The blood-test procedure at a clinic is that a person arrives, takes a numbered ticket and waits for that number to be called. The waiting times between the numbers called have independent normal distributions with mean 3.5 minutes and standard deviation 0.9 minutes. My ticket is number 39 and as I take my ticket number 1 is being called, so that I have to wait for 38 numbers to be called. Find the probability that I will have to wait between 120 minutes and 140 minutes. **[6]**

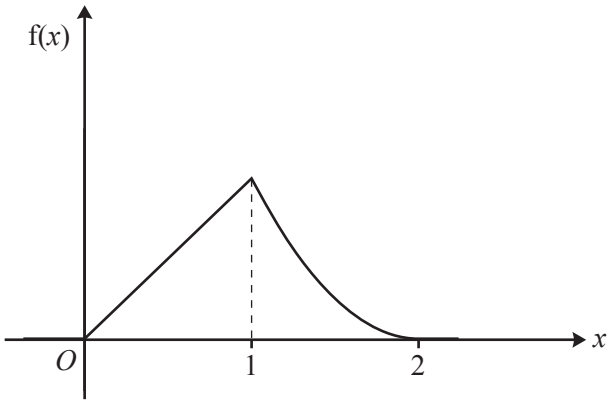
2 In order to estimate the total number of rabbits in a certain region, a random sample of 500 rabbits is captured, marked and released. After two days a random sample of 250 rabbits is captured and 24 are found to be marked. It may be assumed that there is no change in the population during the two days.

(i) Estimate the total number of rabbits in the region. **[2]**

(ii) Calculate an approximate 95% confidence interval for the population proportion of marked rabbits. **[4]**

(iii) Using your answer to part **(ii)**, estimate a 95% confidence interval for the total number of rabbits in the region. **[3]**

3



The continuous random variable X has probability density function given by

$$\mathrm{f}(x) = \begin{cases} ax & 0 < x \leqslant 1, \\ b(2-x)^2 & 1 < x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where a and b are constants. The graph is shown in the above diagram.

(i) Find the values of a and b. **[5]**

(ii) Find the value of $\mathrm{E}\left(\frac{1}{X}\right)$. **[3]**

4 A new computer was bought by a local council to search council records and was tested by an employee. She searched a random sample of 500 records and the sample mean search time was found to be 2.18 milliseconds and an unbiased estimate of variance was 1.58^2 milliseconds^2.

(i) Calculate a 98% confidence interval for the population mean search time μ milliseconds. **[3]**

(ii) It is required to obtain a sample mean time that differs from μ by less than 0.05 milliseconds with probability 0.95. Estimate the sample size required. **[4]**

(iii) State why it is unnecessary for the validity of your calculations that search time has a normal distribution. **[1]**

5 The continuous random variable Y has probability density function given by

$$f(y) = \begin{cases} \ln(y) & 1 \leqslant y \leqslant e, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Verify that this is a valid probability density function. **[4]**

(ii) Show that the (cumulative) distribution function of Y is given by

$$F(y) = \begin{cases} 0 & y < 1, \\ y\ln y - y + 1 & 1 \leqslant y \leqslant e, \\ 1 & \text{otherwise.} \end{cases}$$ **[3]**

(iii) Verify that the upper quartile of Y lies in the interval [2.45, 2.46]. **[2]**

(iv) Find the (cumulative) distribution function of X where $X = \ln Y$. **[4]**

6 A random sample of 80 students who had all studied Biology, Chemistry and Art at a college was each asked which they enjoyed most. The results, classified according to gender, are given in the table.

		Subject		
		Biology	Chemistry	Art
Gender	Male	13	4	11
	Female	37	8	7

It is required to carry out a test of independence between subject most enjoyed and gender at the $2\frac{1}{2}\%$ significance level.

(i) Calculate the expected values for the cells. **[3]**

(ii) Explain why it is necessary to combine cells, and choose a suitable combination. **[2]**

(iii) Carry out the test. **[8]**

7 Two machines A and B both pack cartons in a factory. The mean packing times are compared by timing the packing of 10 randomly chosen cartons from machine A and 8 randomly chosen cartons from machine B. The times, t seconds, taken to pack these cartons are summarised below.

	Sample size	Σt	Σt^2
Machine A	10	221.4	4920.9
Machine B	8	199.2	4980.3

The packing times have independent normal distributions.

(i) Stating a necessary assumption, carry out a test, at the 1% significance level, of whether the population mean packing times differ for the two machines. **[12]**

(ii) Find the largest possible value of the constant c for which there is evidence at the 1% significance level that $\mu_B - \mu_A > c$, where μ_B and μ_A denote the respective population mean packing times in seconds. **[3]**

Question			Answer	Marks	Guidance
1			Total time $T \sim N(\mu, \sigma^2)$	M1	Using $\sum T_i \sim N$
			$\mu = 38 \times 3.5$	A1	= 133 129.5 (from 37) or 136.5 (from 39) A0
			$\sigma^2 = 38 \times 0.9^2$	A1	= 30.78 29.97 (from 37) or 31.59 (from 39) A0
			$P(120 < T < 140) =$		M1 for standardising and combining . Allow even if spurious cc
			$\Phi[(140 - 133)/\sigma] - \Phi[(120 - 133)/\sigma]$	M1	or σ^2 or from $38^2 \times 0.9^2$ used.
			$= 0.8966 - 0.0095$	A1	allow 0.9724 – 0.0414 (from 37) or 0.7336 – 0.0017 (from 39) A1ft
			$= 0.887$	A1	= 0.931 or = 0.732 A1ft
				[6]	
2	(i)		$24/250 = 500/n$	M1	AEF
			$n = 500 \times 250/24 \approx 5208$	A1	Or 5210 or 5200. Must be integer.
				[2]	
2	(ii)		$p_s = 24/250$ (0.096)		
			$s^2 = 0.096 \times 0.904/250 = 0.000347$ or $s = 0.0186$	B1	or /249
			Use $p_s \pm zs$	M1	$p_s + s$ M0. Incorrect n, eg 5208, B0M1B1A0
			$z = 1.96$	B1	
			(0.05948, 0.13252) ~ (0.0595, 0.133)	A1	or (0.0594, 0.1326) Allow 3DP answers. Must be interval, indicated somewhere. Allow limits reversed.
				[4]	
2	(iii)		$500/n_1 =$ "0.0595" and $500/n_2 =$ "0.1325"	M1	
			$n_1 = 8406$ or 8410 (3SF) [8400, 8420]	A1	
			$n_2 = 3774$ or 3770 (3SF) [3759, 3774]	A1	Allow M1A1A0 if both values in range, but neither of them are integers.
				[3]	
3	(i)		(Continuity at $x = 1$ gives) $a = b$	B1	$a = b$ seen or implied, eg by $a = 1.2$, $b = 1.2$
			$\int_0^1 ax\,dx + \int_1^2 b(2-x)^2\,dx = 1$	M1	even from two sep. areas eg 0.5, 0.5.
			$\frac{a}{2}, \left[\frac{-b}{3}(2-x)^3\right]_1^2$ (= 1) oe both seen	B1	Allow $\left[\frac{b}{3}(2-x)^3\right]_1^2$ Allow without limits …+ $b\left[4x - 2x^2 + \frac{x^3}{3}\right]_1^2$.
			$a/2 + b/3 = 1$	B1	cwo
			solving gives $a = b = 6/5$	A1	
				[5]	

Question			Answer	Marks	Guidance
3	(ii)		$\int_0^1 \frac{6}{5}dx + \int_1^2 \frac{6}{5}(\frac{4}{x} - 4 + x)dx$ aef. $= [6x/5 + [(6/5)(4\ln x - 4x + x^2/2)]$ $= (24/5)\ln 2 - 9/5$ AEF or 1.53 (3SF)	M1 A1ft A1 **[3]**	From $\int \frac{f(x)}{x}dx$ ft a, b for M1A1 allow $a + b(4\ln 2 - \frac{5}{2})$. Either their a/b or letters.
4	(i)		Use $2.18 \pm z(1.58/\sqrt{(500)})$ $z = 2.326$ (2.02, 2.34) seconds	M1 B1 A1 **[3]**	allow t Interval seen somewhere, allow reversed.
4	(ii)		Requires $P(\lvert \bar{x} - \mu \rvert < 0.05) = 0.95$ $P(z < 0.05/(1.58/\sqrt{n})) = 0.95$ $0.05\sqrt{n}/1.58 > 1.96$ or $= 1.96$ n = [3836, 3840]	M1 M1 A1 A1 **[4]**	Setting up an equality or equation, even if not z. Standardising. Expression not involving 2.18. Can have × 2. Allow use of same incorrect sd as part (i) for 2nd M1 All but n numeric NOT < or >. Integer required.
4	(iii)		n large enough for sample mean to approximate to normal	B1 **[1]**	Must be mean. Allow n large enough for CLT to apply. NOT just 'n large'.
5	(i)		f is non-negative over [1, e] Attempt to show area, between 1 and e, = 1 $\int_1^e \ln y dy = y\ln y - \int dy$ $= [y\ln y - y]_1^e$ $= (e\ln e - e) - (1\ln 1 - 1) = 1$	B1 M1 M1 A1 **[4]**	 Integrate by parts. Allow 1 error. cwo
5	(ii)		$F(y) = \int_1^y \ln t dt = [t\ln t - t]_1^y$ $= (y\ln y - y) - (1\ln 1 - 1)$ $= y\ln y - y + 1$ AG, over [1, e]	M1 A1 A1 **[3]**	Limits must be correct. If indef integral, must have '+c' and attempt to evaluate. $1\ln 1 - 1 + c = 0$ OR $e\ln e - e + c = 1$ Needs proper justification.
5	(iii)		F(2.45) = 0.745, F(2.46) = 0.754 and 0.745 < 0.75 < 0.754 and result follows	M1 A1 **[2]**	or 1 – F = 0.255, 0.246 or $y\ln y - y + 1 = 0.75$ oe

Question			Answer	Marks	Guidance
5	(iv)		$G(x) = P(X < x)$ $= P(\ln Y < x)$ $= P(Y < e^x)$ $= F(e^x)$ $= (e^x \ln e^x - e^x + 1)$ $xe^x - e^x + 1$; over [0,1]	 M1 M1 A1;B1 **[4]**	 Allow M0M1A1B1.
6	(i)		17.5 4.2 6.3 32.5 7.8 11.7 oe	M1 A1 A1 **[3]**	eg $50 \times 28 \div 80$ At least 2 correct. All correct.
6	(ii)		The E value of $4.2 < 5$ Combine Biology and Chemistry (both sciences).	B1 B1 **[2]**	Need not mention 4.2 May need to look at (iii) to see which subjects combined.
6	(iii)		Ho: Subject and sex are independent H_1: They are not independent 21.7 6.3 40.3 11.7 $\chi^2 = (4.7 - 0.5)^2(21.7^{-1} + 6.3^{-1} + 40.3^{-1} + 11.7^{-1})$ $= 5.558\ldots$ ($v = 1$) (α) 2½ % CV = 5.024 5.558 > CV or in CR and reject H_0	 B1 B1 M1M1 A1 B1 M1	 oe. NOT 'variables', 'they' etc or 17.5 10.5 32.5 19.5 if C/A combined. No Yates(inc $\upsilon > 1$) or incorrect Yates (eg no modulus) M1M0. allow 6.96 or 6.79 Chem./Art combined B1B1M1M1A0B1M1A0. (TS = 3.75) ft TS & CV. Correct first conclusion. If C/A prob. accept H_0.
			(β) $P(\chi^2_1 \geq 5.558) = 0.0184$ < 0.025 and reject H_0	B1 M1	
			There is significant evidence that subject and sex are not independent	A1 **[8]**	cwo. NOT over-assertive. Thus no or incorrect Yates can score max 6/8 B1B1M1M0A1B1M1A0.

Question			Answer	Marks	Guidance
7	(i)		(Assumes the populations of machine)times have equal variances	B1	unless explicitly samples.
			$H_0: \mu_A = \mu_B$, $H_1: \mu_A \neq \mu_B$	B1	For both. Accept in words, provided 'population' used.
			Pooled estimate = s^2		
			= [4920.9 – 221.4²/10 + 4980.3 – 199.2²/8]/16 = 39.324/16	M1A1	A1 for 39.324 or 16. Not pooled
			= 2.45775	A1	M0A0A0 here,
			SD of $\overline{t}_B - \overline{t}_A = s\sqrt{(10^{-1} + 8^{-1})}$		
			= 0.74364	A1	A0 for 0.757
			1% CV = 2.921	B1	allow B1 for 2.921
			$\overline{t}_B - \overline{t}_A = 2.76$ or 24.9 – 22.14	B1	allow B1 for 2.76 etc
			Test statistic T = 2.76/SD	M1	Needs denominator (s) $(\frac{\cdot\cdot}{10}, \frac{\cdot\cdot}{8}), (\frac{\cdot\cdot}{9}, \frac{\cdot\cdot}{7}), \frac{\cdot\cdot}{18}, \frac{\cdot\cdot}{16}$ for M1.
			= 3.711	A1	allow A1ft for 3.645
			T > 2.921 so reject H_0.	M1	ft TS,CV even if 1-tail and/or z. Consistent signs.
			There is sufficient evidence that the population mean packing times differ	A1	ft TS only . 'Not pooled' can score B1B1M0A0A0A0B1B1M1A1M1A1 max 8/12
				[12]	
7	(ii)		CV 2.583	B1	
			$(2.76 - c)/\text{SD} \geq \text{CV}$	M1	Allow =, ≤; must be sd from (i).
			$c \leq 0.8392$ or largest ≈ 0.839	A1	NOT isw.
				[3]	