OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS

4734

Probability & Statistics 3

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- List of Formulae (MF1)

Other Materials Required:
None

Friday 19 June 2009
Afternoon

Duration: 1 hour 30 minutes

* 4 7 3 4 *

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- 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.
- You are permitted to use a graphical calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets **[ ]** at the end of each question or part question.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- This document consists of **4** pages. Any blank pages are indicated.

1 A continuous random variable X has probability density function given by

$$\mathrm{f}(x) = \begin{cases} \dfrac{2x}{5} & 0 \leqslant x \leqslant 1, \\ \dfrac{2}{5\sqrt{x}} & 1 < x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

Find

(i) $\mathrm{E}(X)$, **[3]**

(ii) $\mathrm{P}\big(X \geqslant \mathrm{E}(X)\big)$. **[3]**

2 The number of bacteria in 1 ml of drug A has a Poisson distribution with mean 0.5. The number of the same bacteria in 1 ml of drug B has a Poisson distribution with mean 0.75. A mixture of these drugs used to treat a particular disease consists of 1.4 ml of drug A and 1.2 ml of drug B. Bacteria in the drugs will cause infection in a patient if 5 or more bacteria are injected.

(i) Calculate the probability that, in a sample of 20 patients treated with the mixture, infection will occur in no more than one patient. **[7]**

(ii) State an assumption required for the validity of the calculation. **[1]**

3 A machine produces circular metal discs whose radii have a normal distribution with mean μ cm. A random sample of five discs is selected and their radii, in cm, are as follows.

6.47 6.52 6.46 6.47 6.51

(i) Calculate a 95% confidence interval for μ. **[6]**

(ii) Hence state a 95% confidence interval for the mean circumference of a disc. **[1]**

4 In order to compare the difficulty of two Su Doku puzzles, two random samples of 40 fans were selected. One sample was given Puzzle 1 and the other sample was given Puzzle 2. Of those given Puzzle 1, 24 could solve it within ten minutes. Of those given Puzzle 2, 15 could solve it within ten minutes.

(i) Using proportions, test at the 5% significance level whether there is a difference in the standard of difficulty of the two puzzles. **[8]**

(ii) The setter believed that Puzzle 2 was more difficult than Puzzle 1. Obtain the smallest significance level at which this belief is supported. **[2]**

5 Each person in a random sample of 15 men and 17 women from a university campus was asked how many days in a month they took exercise. The numbers of days for men and women, x_M and x_W respectively, are summarised by

$$\Sigma x_M = 221, \quad \Sigma x_M^2 = 3992, \quad \Sigma x_W = 276, \quad \Sigma x_W^2 = 5538.$$

(i) State conditions for the validity of a suitable test of the difference in the mean numbers of days for men and women on the campus. **[3]**

(ii) Given that these conditions hold, carry out the test at the 5% significance level. **[10]**

(iii) If in fact the random sample was drawn entirely from the university Mathematics Department, state with a reason whether the validity of the test is in doubt. **[1]**

6 The function $\mathrm{F}(t)$ is defined as follows.

$$\mathrm{F}(t) = \begin{cases} 0 & t < 0, \\ \sin^4 t & 0 \leqslant t \leqslant \frac{1}{2}\pi, \\ 1 & t > \frac{1}{2}\pi. \end{cases}$$

(i) Verify that F is a (cumulative) distribution function. **[2]**

The continuous random variable T has (cumulative) distribution function $\mathrm{F}(t)$.

(ii) Find the lower quartile of T. **[3]**

(iii) Find the (cumulative) distribution function of Y, where $Y = \sin T$, and obtain the probability density function of Y. **[5]**

(iv) Find the expected value of $\dfrac{1}{Y^3 + 2Y^4}$. **[3]**

7 In 1761, James Short took measurements of the parallax of the sun based on the transit of Venus. The mean and standard deviation of a random sample of 50 of these measurements are 8.592 and 0.7534 respectively, in suitable units.

(i) Show that if $X \sim \mathrm{N}(8.592, 0.7534^2)$, then

$$\mathrm{P}(X \leqslant 8.084) = \mathrm{P}(8.084 < X \leqslant 8.592) = \mathrm{P}(8.592 < X \leqslant 9.100) = \mathrm{P}(X > 9.100) = 0.25.$$ **[4]**

The following table summarises the 50 measurements using these intervals.

Measurement (x)	$x \leqslant 8.084$	$8.084 < x \leqslant 8.592$	$8.592 < x \leqslant 9.100$	$x > 9.100$
Frequency	8	22	11	9

(ii) Carry out a test, at the $\frac{1}{2}$% significance level, of whether a normal distribution fits the data. **[7]**

(iii) Obtain a 99% confidence interval for the mean of all similar parallax measurements. **[3]**

4734 Probability & Statistics 3

Penalise 2 sf instead of 3 once only. Penalise final answer ≥ 6 sf once only.

1 (i)	$\int_0^1 \frac{2}{5}x^2 dx + \int_1^4 \frac{2}{5}\sqrt{x}\,dx$	M1		Attempt to integrate xf(x), both parts added, limits
	$= \left[\frac{2x^3}{15}\right]_0^1 + \left[\frac{4x^{3/2}}{15}\right]_1^4 = 2$	A1		Correct indefinite integrals
		A1	3	Correct answer
(ii)	$\int_2^4 \frac{2}{5\sqrt{x}} dx = \left[\frac{4\sqrt{x}}{5}\right]_2^4 = \frac{4}{5}(2-\sqrt{2})$ or 0.4686	M1		Attempt correct integral, limits; needs "1 –" if $\mu < 1$
		A1		Correct indefinite integral, √ on their μ
		A1	3	Exact aef, or in range [0.468, 0.469]
2 (i)	Po(0.5), Po(0.75)	M1		0.5, 0.75 scaled
	Po(0.7) and Po(0.9)	A1		These
	$A + B \sim$ Po(1.6)	M1		Sum of Poissons used, can have wrong parameters
	P($A + B \geq 5$) = 0.0237	A1		0.0237 from tables or calculator
	B(20, 0.0237)	M1		Binomial (20, their p), soi
	$0.9763^{20} + 20\times0.9763^{19}\times0.0237$	A1√		Correct expression, their p
	= 0.9195	A1	7	Answer in range [0.919, 0.92]
(ii)	Bacteria should be independent in drugs; *or* sample should be random	B1	1	Any valid relevant comment, must be contextualised
3 (i)	Sample mean = 6.486	B1		
	s^2 = 0.00073	B1		0.000584 if divided by 5
	$6.486 \pm 2.776 \times \sqrt{\frac{0.00073}{5}}$	M1		Calculate sample mean ± $ts/\sqrt{5}$, allow 1.96, s^2 etc
		B1		t = 2.776 seen
	(6.45, 6.52)	A1A1	6	Each answer, cwo (6.45246, 6.5195)
(ii)	2π × above [= (40.5, 41.0)]	M1	1	
4 (i)	H_0: $p_1 = p_2$; H_1: $p_1 \neq p_2$, where p_i is the proportion of all solvers of puzzle i	B1		Both hypotheses correctly stated, allow eg $\hat{p}$
	Common proportion 39/80	M1A1		[= 0.4875]
	$s^2 = 0.4875 \times 0.5125 / 20$	B1		[= 0.01249, σ = 0.11176]
	$(\pm)\frac{0.6-0.375}{0.1117} = (\pm)2.013$	M1		$(0.6 - 0.375)/s$
		A1√		Allow 2.066√ from unpooled variance, p = 0.0195
	2.013 > 1.96, or 0.022 < 0.025	M1		Correct method and comparison with 1.96 or 0.025, allow unpooled, 1.645 from 1-tailed *only*
	Reject H_0. Significant evidence that there is a difference in standard of difficulty	A1√	8	Conclusion, contextualised, not too assertive
(ii)	One-tail test used	M1		One-tailed test stated or implied by
	Smallest significance level 2.2(1)%	A1	2	Φ("2.013"), OK if off-scale; allow 0.022(1)

5 (i)	Numbers of men and women should have normal dists; with equal variance; distributions should be independent	B1 B1 B1 **3**	Context & 3 points: 2 of these, B1; 3, B2; 4, B3. [Summary data: 14.73 49.06 52.57 16.24 62.18 66.07]
(ii)	$H_0: \mu_M = \mu_W$; $H_1: \mu_M \neq \mu_W$ $3992 - \frac{221^2}{15} + 5538 - \frac{276^2}{17}$ [≈ 1793] 1793/(14 + 16) = 59.766 $(\pm)\frac{221/15 - 276/17}{\sqrt{59.766(\frac{1}{15}+\frac{1}{17})}} = (-)0.548$ Critical region: $\lvert t \rvert \geq 2.042$ Do not reject H_0. Insufficient evidence of a difference in mean number of days	B1 M1 A1 A1 M1 A1√ A1 B1 M1 A1√ **10**	Both hypotheses correctly stated Attempt at this expression (see above) Either 1793 or 30 Variance estimate in range [59.7, 59.8] (or √ = 7.73) Standardise, allow wrong (but not missing) $1/n$ Correct formula, allow $s^2(\frac{1}{15}+\frac{1}{17})$ or $(\frac{s_1^2}{15}+\frac{s_2^2}{17})$, allow 14 & 16 in place of 15, 17; 0.548 or –0.548 2.042 seen Correct method and comparison type, must be t, allow 1-tail; conclusion, in context, not too assertive
(iii)	Eg Samples not indep't so test invalid	B1 **1**	Any relevant valid comment, eg "not representative"

6 (i)	F(0) = 0, F(π/2) = 1 Increasing	B1 B1 **2**	Consider both end-points Consider F between end-points, can be asserted
(ii)	$\sin^4(Q_1) = ¼$ $\sin(Q_1) = 1/\sqrt{2}$ $Q_1 = \pi/4$	M1 A1 A1 **3**	 Can be implied. Allow decimal approximations Or 0.785(4)
(iii)	G(y) $= P(Y \le y) = P(T \le \sin^{-1} y)$ $= F(\sin^{-1} y)$ $= y^4$ $g(y) = \begin{cases} 4y^3 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$	M1 A1 A1 M1 A1 **5**	 Ignore other ranges Differentiate G(y) Function and range stated, allow if range given in G
(iv)	$\int_0^1 \frac{4}{1+2y}\,dy = [2\ln(1+2y)]_0^1$ = **2 ln 3**	M1 A1 A1 **3**	Attempt $\int \frac{g(y)}{y^3+2y^4}\,dy$; $\int_0^1 \frac{4}{1+2y}\,dy$ Or 2.2, 2.197 or better
7 (i) α	$\Phi\left(\frac{8.084-8.592}{0.7534}\right) = \Phi(-0.674) = 0.25$ $\Phi(0) - \Phi(\text{above}) = 0.25$ $P(8.592 \le X \le 9.1)$ = same by symmetry	M1 A1 A1 A1 **4**	Standardise once, allow $\sqrt{}$ confusions, ignore sign Obtain 0.25 for one interval For a second interval, justified, eg using $\Phi(0) = 0.5$ For a third, justified, eg "by symmetry"
or β	$\frac{x-8.592}{0.7534} = 0.674$ $x = 8.592 \pm 0.674 \times 0.7534$ = (8.084, 9.100)	M1A1 A1A1	[from probabilities to ranges] A1 for art 0.674
(ii)	H_0: normal distribution fits data All E values 50/4 = 12.5 $X^2 = \frac{4.5^2+9.5^2+1.5^2+3.5^2}{12.5} = 10$ 10 > 7.8794 Reject H_0. Significant evidence that normal distribution is not a good fit.	B1 B1 M1 A1 B1 M1 A1$\sqrt{}$ **7**	*Not* N(8.592, 0.7534). Allow "it's normally distributed" [Yates: 8.56: A0] CV 7.8794 seen Correct method, incl. formula for χ^2 and comparison, allow wrong ν Conclusion, in context, not too assertive
(iv)	$8.592 \pm 2.576 \times \frac{0.7534}{\sqrt{49}}$ (8.315, 8.869)	M1 A1 A1 **3**	Allow $\sqrt{}$ errors, wrong σ or z, allow 50 Correct, including $z = 2.576$ or $t_{49} = 2.680$, *not* 50 In range [8.31, 8.32] and in range (8.86, 8.87], even from 50, or (8.306, 8.878) from t_{49}