OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 4733

Probability & Statistics 2

Specimen Paper

Additional materials:
Answer booklet
Graph paper
List of Formulae (MF 1)

TIME 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your Name, Centre Number and Candidate Number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- 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 graphic calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- The total number of marks for this paper is 72.
- Questions carrying smaller numbers of marks are printed earlier in the paper, and questions carrying larger numbers of marks later in the paper.
- **You are reminded of the need for clear presentation in your answers.**

This question paper consists of 3 printed pages and 1 blank page.

Registered Charity Number: 1066969
[Turn over

1 The standard deviation of a random variable F is 12.0. The mean of n independent observations of F is denoted by $\overline{F}$.

(i) Given that the standard deviation of $\overline{F}$ is 1.50, find the value of n. [3]

(ii) For this value of n, state, with justification, what can be said about the distribution of $\overline{F}$. [2]

2 A certain neighbourhood contains many small houses (with small gardens) and a few large houses (with large gardens). A sample survey of all houses is to be carried out in this neighbourhood. A student suggests that the sample could be selected by sticking a pin into a map of the neighbourhood the requisite number of times, while blindfolded.

(i) Give two reasons why this method does not produce a random sample. [2]

(ii) Describe a better method. [3]

3 Sixty people each make two throws with a fair six-sided die.

(i) State the probability of one particular person obtaining two sixes. [1]

(ii) Using a suitable approximation, calculate the probability that at least four of the sixty obtain two sixes. [5]

4 The random variable G has mean 20.0 and standard deviation σ. It is given that $\mathrm{P}(G > 15.0) = 0.6$. Assume that G is normally distributed.

(i) **(a)** Find the value of σ. [4]

(b) Given that $\mathrm{P}(G > g) = 0.4$, find the value of $\mathrm{P}(G > 2g)$. [3]

(ii) It is known that no values of G are ever negative. State with a reason what this tells you about the assumption that G is normally distributed. [2]

5 The mean solubility rating of widgets inserted into beer cans is thought to be 84.0, in appropriate units. A random sample of 50 widgets is taken. The solubility ratings, x, are summarised by

$$n = 50, \qquad \Sigma x = 4070, \qquad \Sigma x^2 = 336\,100 .$$

Test, at the 5% significance level, whether the mean solubility rating is less than 84.0. [10]

6 On average a motorway police force records one car that has run out of petrol every two days.

(i) **(a)** Using a Poisson distribution, calculate the probability that, in one randomly chosen day, the police force records exactly two cars that have run out of petrol. [3]

(b) Using a Poisson distribution and a suitable approximation to the binomial distribution, calculate the probability that, in one year of 365 days, there are fewer than 205 days on which the police force records no cars that have run out of petrol. [6]

(ii) State an assumption needed for the Poisson distribution to be appropriate in part **(i)**, and explain why this assumption is unlikely to be valid. [2]

7 The time, in minutes, for which a customer is prepared to wait on a telephone complaints line is modelled by the random variable X. The probability density function of X is given by

$$\mathrm{f}(x) = \begin{cases} kx(9 - x^2) & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{4}{81}$. [2]

(ii) Find $\mathrm{E}(X)$. [3]

(iii) **(a)** Show that the value y which satisfies $\mathrm{P}(X < y) = \frac{3}{5}$ satisfies

$$5y^4 - 90y^2 + 243 = 0.$$ [4]

(b) Using the substitution $w = y^2$, or otherwise, solve the equation in part **(a)** to find the value of y. [3]

8 The proportion of left-handed adults in a country is known to be 15%. It is suggested that for mathematicians the proportion is greater than 15%. A random sample of 12 members of a university mathematics department is taken, and it is found to include five who are left-handed.

(i) Stating your hypotheses, test whether the suggestion is justified, using a significance level as close to 5% as possible. [8]

(ii) In fact the significance test cannot be carried out at a significance level of exactly 5%. State the probability of making a Type I error in the test. [2]

(iii) Find the probability of making a Type II error in the test for the case when the proportion of mathematicians who are left-handed is actually 20%. [2]

(iv) Determine, as accurately as the tables of cumulative binomial probabilities allow, the actual proportion of mathematicians who are left-handed for which the probability of making a Type II error in the test is 0.01. [2]

1	**(i)**	$\frac{12.0}{\sqrt{n}} = 1.50 \Rightarrow \sqrt{n} = \frac{12.0}{1.50} = 8 \Rightarrow n = 64$	B1	For any correct equation involving n
			M1	For correct solution method for n or $\sqrt{n}$
			A1 **3**	For correct answer 64
	(ii)	n is large, the distribution of $\bar{F}$ can be taken to be normal, according to the Central Limit Theorem	M1	For relating the size of n to normality
			A1 **2**	For reference to the CLT
			5	
2	**(i)**	Reasons for bias may include:		
		Larger properties more likely to be picked	B1	For stating one valid relevant reason
		Some regions of the map more/less likely	B1 **2**	For stating a second valid relevant reason
	(ii)	Make a list of all the houses in the neighbourhood	B1	For stating or implying a sampling frame
		Number the houses from 1 upwards	B1	For numbering the sampling units
		Select the sample using random numbers	B1 **3**	For referring to use of random numbers
			5	
3	**(i)**	$\frac{1}{36}$	B1 **1**	For correct probability
	(ii)	Number obtaining two sixes ~ B(60, $\frac{1}{36}$)	M1	For stating or implying binomial distribution
		Approximate distribution is Po($\frac{5}{3}$)	A1√	For the correct Poisson approximation
		$P(\geqslant 4) = 1 - e^{-\frac{5}{3}}\left\{1 + \frac{5}{3} + \frac{(5/3)^2}{2!} + \frac{(5/3)^3}{3!}\right\}$	M1	For calculation of correct terms
			M1	For correct use of Poisson formula
		$= 0.0883$	A1 **5**	For correct answer 0.088(3)
			6	
4	**(i)**	**(a)** $\frac{15.0 - 20.0}{\sigma} = -0.253$	M1	For standardising and equating to $\Phi^{-1}(p)$
			B1	For correct value 0.253 (or 0.254) seen
		Hence $\sigma = \frac{5}{0.253} \approx 19.8$	M1	For solving equation for σ
			A1 **4**	For correct value 19.8
		(b) $g = 25.0$, using symmetry	B1	For stating (or finding) the value of g
		Hence $P(G > 2g) = 1 - \Phi\left(\frac{50.0 - 20.0}{19.8}\right)$	M1	For correct process for upper tail prob
		$= 1 - 0.935 = 0.065$	A1 **3**	For correct answer
	(ii)	If normal, $P(G < 0)$ is substantial	M1	For considering relevant normal probability
		Hence the assumption seems unjustified	A1 **2**	For stating the appropriate conclusion
			9	

5	$\bar{x} = \frac{4070}{50} = 81.4$	B1	For correct value of sample mean
	$s^2 = \frac{336\,100}{49} - \frac{4070^2}{49 \times 50} = 98$	M1	For calculation of unbiased or biased estimate
		A1	For correct value of unbiased estimate
	$H_0 : \mu = 84.0; \quad H_1 : \mu < 84.0$	B1	For correct statement of null hypothesis
		B1	For correct statement of alt hypothesis
	EITHER: $z = \frac{\bar{x} - 84.0}{\sqrt{(s^2/50)}} = -1.857$	M1	For standardising, including use of $\sqrt{50}$
		A1	For correct value 1.857
	This is significant, since $-1.857 < -1.645$	M1	For comparing z value to -1.645 or equiv
	OR: $\frac{c - 84.0}{\sqrt{(s^2/50)}} = -1.645 \Rightarrow c = 81.697$	M1	For critical value calculation, inc use of $\sqrt{50}$
		A1	For correct value 81.697
	$\bar{x}$ is in the critical region since $81.4 < 81.697$	M1	For comparing sample mean to critical region
	Hence H_0 is rejected	A1√	For stating or implying rejection of H_0
	There is sufficient evidence to conclude that the mean solubility rating is less than 84.0	A1√ **10**	For stating the outcome in context
		10	
6	**(i)** **(a)** For one day, the distribution is Po(0.5)	B1	For use of correct Poisson mean
	Hence P(exactly 2) $= 0.9856 - 0.9098$	M1	For relevant use of tables (or formula)
	$= 0.0758$	A1 **3**	For correct answer 0.0758
	(b) No. of days with no cars ~ B(365, 0.6065)	M1	For relevant Poisson probability of P(0)
		A1	For identifying correct binomial distribution
	Normal approximation is N(221.3725, 87.11)	A1√	For correct use of *np* and *npq*
	$P(< 205) = P\left(Z < \frac{204.5 - 221.3725}{\sqrt{87.11}}\right)$	M1	For standardising (with or without c.c. here)
		A1	For completely correct expression
	$= \Phi(-1.808) = 0.0353$	A1 **6**	For correct answer 0.0353
	(ii) Events (cars running out of petrol) must occur at a constant average rate. This seems unlikely, given that there will be different volumes of traffic on different days of the week (e.g. weekdays and weekends)	B1	For correct statement of the condition
		B1 **2**	For a correct explanation
		11	

[Turn over

7	**(i)**	$1 = k\int_0^3 (9x - x^3)\,\mathrm{d}x = k\left[\frac{9}{2}x^2 - \frac{1}{4}x^4\right]_0^3 = \frac{81}{4}k$	M1	For equating to 1 and integrating
		Hence $k = \frac{4}{81}$	A1 **2**	For showing given answer correctly
	(ii)	$\mathrm{E}(X) = \frac{4}{81}\int_0^3 x^2(9 - x^2)\,\mathrm{d}x = \frac{4}{81}\left[3x^3 - \frac{1}{5}x^5\right]_0^3 = 1.6$	M1	For attempt at $\int_0^3 x\,\mathrm{f}(x)\,\mathrm{d}x$
			A1	For correct indefinite integral, in any form
			A1 **3**	For correct answer 1.6
	(iii)	**(a)** $\frac{3}{5} = \frac{4}{81}\int_0^y x(9 - x^2)\,\mathrm{d}x = \frac{4}{81}\left[\frac{9}{2}x^2 - \frac{1}{4}x^4\right]_0^y$	M1	For attempt at $\int_0^y \mathrm{f}(x)\,\mathrm{d}x = \frac{3}{5}$
			B1	For correct indefinite integral, in any form
		Hence $\frac{3}{5} = \frac{4}{81}\left\{\frac{9}{2}y^2 - \frac{1}{4}y^4\right\}$	M1	Use limits to produce relevant equation in y
		i.e. $5y^4 - 90y^2 + 243 = 0$	A1 **4**	For showing given answer correctly
		(b) $w = \dfrac{90 \pm \sqrt{(90^2 - 4\times5\times243)}}{10} = 3.31$ or 14.7	M1	For use of quadratic formula to find w
			A1	For either value found correctly
		Hence $y = \sqrt{3.31} = 1.82$	A1 **3**	For correct (unique) answer 1.82
			12	
8	**(i)**	$\mathrm{H}_0 : p = 0.15$; $\mathrm{H}_1 : p > 0.15$	B1	For correct statement of null hypothesis
			B1	For correct statement of alt hypothesis
		Under H_0, number left-handed $L \sim \mathrm{B}(12, 0.15)$	M1	For correct distribution stated or implied
		$\mathrm{P}(L \geqslant 5) = 1 - 0.9761 = 0.0239$	M1	For calculation of relevant tail probability, or finding the critical region
			A1	For correct value 0.0239 or region $l \geqslant 5$
		This is significant, since $0.0239 < 0.05$	M1	For comparing tail probability with 0.05 or observed value with critical region
		Hence H_0 is rejected	A1√	For stating or implying rejection of H_0
		Accept the suggestion that the proportion of mathematicians who are left-handed is more than 15%	A1√ **8**	For stating the outcome in context
	(ii)	$\mathrm{P_I} = \mathrm{P}(L$ in critical region$) = 0.0239$	M1	For evaluating $\mathrm{P}(\text{reject } \mathrm{H}_0)$
			A1 **2**	For correct answer 0.0239 or equivalent
	(iii)	$\mathrm{P_{II}} = \mathrm{P}(L \leqslant 4 \mid p = 0.2) = 0.9274$	M1	For evaluating $\mathrm{P}(\text{accept } \mathrm{H}_0)$ with $p = 0.2$
			A1 **2**	For correct probability
	(iv)	$\mathrm{P_{II}} = 0.0188$ for $p = \frac{2}{3}$ and 0.0095 for $p = 0.7$	M1	For relevant use of tables
		So the proportion is between 67% and 70%	A1 **2**	For an appropriate conclusion
			14	