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

Thursday 15 January 2009
Morning

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 At a particular hospital, admissions of patients as a result of visits to the Accident and Emergency Department occur randomly at a uniform average rate of 0.75 per day. Independently, admissions that result from G.P. referrals occur randomly at a uniform average rate of 6.4 per *week*. The total number of admissions from these two causes over a randomly chosen period of four weeks is denoted by T. State the distribution of T and obtain its expectation and variance. **[4]**

2 The continuous random variable U has (cumulative) distribution function given by

$$\mathrm{F}(u) = \begin{cases} \frac{1}{5}\mathrm{e}^{u} & u < 0, \\ 1 - \frac{4}{5}\mathrm{e}^{-\frac{1}{4}u} & u \geqslant 0. \end{cases}$$

(i) Find the upper quartile of U. **[3]**

(ii) Find the probability density function of U. **[2]**

3 In a random sample of credit card holders, it was found that 28% of them used their card for internet purchases.

(i) Given that the sample size is 1200, find a 98% confidence interval for the percentage of all credit card holders who use their card for internet purchases. **[4]**

(ii) Estimate the smallest sample size for which a 98% confidence interval would have a width of at most 5%, and state why the value found is only an estimate. **[4]**

4 The weekly sales of petrol, X thousand litres, at a garage may be modelled by a continuous random variable with probability density function given by

$$\mathrm{f}(x) = \begin{cases} c & 25 \leqslant x \leqslant 45, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant. The weekly profit, in £, is given by $(400\sqrt{X} - 240)$.

(i) Obtain the value of c. **[1]**

(ii) Find the expected weekly profit. **[3]**

(iii) Find the probability that the weekly profit exceeds £2000. **[3]**

5 The concentration level of mercury in a large lake is known to have a normal distribution with standard deviation 0.24 in suitable units. At the beginning of June 2008, the mercury level was measured at five randomly chosen places on the lake, and the sample mean is denoted by $\bar{x}_1$. Towards the end of June 2008 there was a spillage in the lake which may have caused the mercury level to rise. Because of this the level was then measured at six randomly chosen points of the lake, and the mean of this sample is denoted by $\bar{x}_2$.

(i) State hypotheses for a test based on the two samples for whether, on average, the level of mercury had increased. Define any parameters that you use. **[2]**

(ii) Find the set of values of $\bar{x}_2 - \bar{x}_1$ for which there would be evidence at the 5% significance level that, on average, the level of mercury had increased. **[4]**

(iii) Given that the average level had actually increased by 0.3 units, find the probability of making a Type II error in your test, and comment on its value. **[4]**

6 A mathematics examination is taken by 29 boys and 26 girls. Experience has shown that the probability that any boy forgets to bring a calculator to the examination is 0.3, and that any girl forgets is 0.2. Whether or not any student forgets to bring a calculator is independent of all other students. The numbers of boys and girls who forget to bring a calculator are denoted by B and G respectively, and $F = B + G$.

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

(ii) Using suitable approximations to the distributions of B and G, which should be justified, find the smallest number of spare calculators that should be available in order to be at least 99% certain that all 55 students will have a calculator. **[8]**

7 A tutor gives a randomly selected group of 8 students an English Literature test, and after a term's further teaching, she gives the group a similar test. The marks for the two tests are given in the table.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
First test	38	27	55	43	32	24	51	46
Second test	37	26	57	43	30	26	54	48

(i) Stating a necessary condition, show by carrying out a suitable t-test, at the 1% significance level, that the marks do not give evidence of an improvement. **[8]**

(ii) The tutor later found that she had marked the second test too severely, and she decided to add a constant amount k to each mark. Find the least integer value of k for which the increased marks would give evidence of improvement at the 1% significance level. **[3]**

[Question 8 is printed overleaf.]

8 A soft drinks factory produces lemonade which is sold in packs of 6 bottles. As part of the factory's quality control, random samples of 75 packs are examined at regular intervals. The number of underfilled bottles in a pack of 6 bottles is denoted by the random variable X. The results of one quality control check are shown in the following table.

Number of underfilled bottles	0	1	2	3
Number of packs	44	20	8	3

A researcher assumes that $X \sim B(3, p)$.

(i) By finding the sample mean, show that an estimate of p is 0.2. **[3]**

(ii) Show that, at the 5% significance level, there is evidence that this binomial distribution does not fit the data. **[10]**

(iii) Another researcher suggests that the goodness of fit test should be for $B(6, p)$. She finds that the corresponding value of χ^2 is 2.74, correct to 3 significant figures. Given that the number of degrees of freedom is the same as in part **(ii)**, state the conclusion of the test at the same significance level. **[1]**

4734 Probability & Statistics 3

1	T has a Poisson distribution $E(T)=28\times0.75+4\times6.4$ $= 46.6$ $Var(T)=46.6$	B1 M1 A1 B1√ **4**	From sum of Poissons Ft $E(T)$ only if Poisson
2 **(i)**	Use $F(Q_3)=0.75$ or $\int_{Q_3}^{\infty}\frac{1}{5}e^{-\frac{1}{4}u}\,du = 0.25$ Solve to obtain $Q_3 = 4.65$ AEF eg 4ln(16/5)	M1 M1A1 **3**	M1 for solving similar eqn A0 for ≥4.65
(ii)	$f(u)=\begin{cases}\frac{1}{5}e^{u} & u<0,\\ \frac{1}{5}e^{-\frac{1}{4}u} & u\ge 0.\end{cases}$	B1 B1 **2**	$u < 0$ unless evidence of ∫ $u \ge 0$
3 **(i)**	Use $28\pm zs$ $z=2.326$ $s^2 = 28\times72/1200$ (25.0, 31.0)	M1 B1 B1 A1 **4**	Accept s=c/√n for M1 Accept 0.28 with corresponding s Or 1199 Accept (25, 31)
(ii)	$2\times2.326\sqrt{(0.28\times0.72/n)} \le 0.05$ AEF Solve to obtain n Smallest $n = 1745$ e.g. Variance is an approximation	M1 M1 A1 B1 **4**	Or = or ≥ Solving similar equn Accept 1746 ,1750 Or normal is approx or Or p only an estimate
4 **(i)**	$c = 1/20$	B1 **1**	
(ii)	$\int_{25}^{45}\frac{400\sqrt{x}-240}{20}\,dx$ $=\left[\frac{40}{3}x^{3/2}-12x\right]$ = 2118(£)	M1 A1 A`1 **3**	Correct indefinite integral 2120 or better than 2118
(iii)	$400\sqrt{X}-240 > 2000$, $X > 31.36$ $P(X > 31.36) = (45 - 31.36)/20$ $= 0.682$	M1 M1 A1 **3**	Or 31.4 cao

5 (i)	H_o: μ_2=μ_1, H_1: $\mu_2 > \mu_1$, where μ_1 and μ_2 are the mean concentrations in the lake before and after the spillage respectively	B1 B1 **2**	For both hypotheses Allow in words if population mean used.
(ii)	$\bar{X}_2 - \bar{X}_1 \geq zs$ z=1.645 s=0.24√(1/5+1/6) ≥ 0.2391	M1 A1 B1 A1 **4**	Accept >, =, <. ≤, *ts* Or >; 0.239
(iii)	$P(\bar{X}_2 - \bar{X}_1 < 0.2391)$ z =[0.2391 – 0.3]/s p=0.3376 This is a large probability for this error	M1 M1 A1 B1 **4**	May be implied ART 0.337 or 0.338 Relevant comment
6 (i)	Use B ~ B(29, 0.3) , G ~ B(26, 0.2) E(F)=29×0.3+26×0.2=13.9 Var(F) = 29×0.3×0.7+26×0.2×0.8=10.25	M1 M1A1 M1A1 **5**	
(ii)	B: np = 8.7, nq=20.3 G: np = 5.2, nq= 20.8 All exceed 5, so normal approximation valid for each F ~ N(13.9, 10.25) (approximately) (Requires P($F \leq n$) = 0.99) [n + 0.5 – 13.9]/√(10.25) ; = 2.326, their 10.25 n = 20.85 Need to have 21 spares available SR Using B(55, 0.2527): B1; M1(N(13.9, 10.39); M1B1M1A0 (Max 5/8)	 B2 M1√ M1B1 A1 M1 A1 **8**	Must check numerically B1 for checking one distribution Use normal. May be implied Standardise M0 if variance has divisors cc Solving similar No cc, lose last A1 (n = 22) Wrong cc, lose A1A1

7 (i)	Requires population of (2nd mark – 1st mark) to be normally distributed H_o: $\mu_d = 0$, H_1:$\mu_d > 0$ $T_2 - T_1$: - 1 - 1 2 0 - 2 2 3 2 $\bar{d} = 0.625$, $s^2 = 3.411$ ($3^{23}/_{56}$ or $^{191}/_{56}$) Use 2.998 EITHER: $t = 0.625/\sqrt{(3.411/8)}$ $= 0.957$ OR: CV(CR), $\bar{d} \geq 2.998\sqrt{3.411/8}$ $= 1.958$ EITHER 0.957<2.998 OR 0.625 < 1.958 Do not reject H_o, there is insufficient evidence of improvement	B1 M1 B1B1 B1 M1 A1 M1 A1 M1 **8**	 M0 if clearly z With comparison and conclusion
(ii)	Use $E(X_2 - X_1 + k) = 0.625 + k$ Requires $(0.625+k) / \sqrt{(3.411/8)} \geq 2.998$ Giving $k \geq 1.33$ Increase each mark by 2	M1 A1√ A1 **3**	 Allow 1.33
8 (i)	Mean= (20+16+9)/75 = 0.6 $3p = 0.6$, $p = 0.2$ AG	M1 A1 A1 **3**	
(ii)	H_0: B(3,p) fits the data (H_1: B(3,p) does not fit the data) Expected values 38.4 28.8 7.2 0.6 Combine last two cells $\chi^2 = 5.6^2/38.4 + 8.8^2/28.8 + 3.2^2/7.8$ $= 4.818$ 4.818 > 3.841 Reject H_0 and conclude that there is insufficient evidence that B(3,p) fits the data.	B1 M1 A1 A1 B1 M1 A1√ A1 B1√ M1 **10**	Or: X~B(3,p) or B(3,0.2) Not 'Data fits model' Use B(3,0.2)×75 At least 2 correct All correct With one correct At least 2 correct Ft E values Accept 4.82 cao ft 4.818 SR1 If cells not combined: B1M1A1A1B0M1A1A0B1(5.991)M1 SR2:E-values rounded :B1M1A1A1 B1M1A1A0(4.865)B1M1
(iii)	2.74 < 3.841, accept H_0 conclude that B(6, p) fits the data	B1 **1**	Accept with no reason if evidence of method in (ii)