OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 4734

Probability & Statistics 3

Wednesday **24 MAY 2006** Afternoon 1 hour 30 minutes

Additional materials:
8 page answer booklet
Graph paper
List of Formulae (MF1)

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 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.
- 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 4 printed pages.

Registered Charity Number: 1066969
[Turn over

1 The numbers of α-particles emitted per minute from two types of source, A and B, have the distributions Po(1.5) and Po(2) respectively. The total number of α-particles emitted over a period of 2 minutes from three sources of type A and two sources of type B, all of which are independent, is denoted by X. Calculate $P(X = 27)$. [4]

2 The manager of a factory with a large number of employees investigated when accidents to employees occurred during 8-hour shifts. An analysis was made of 600 randomly chosen accidents that occurred over a year. The following table shows the numbers of accidents occurring in the four consecutive 2-hour periods of the 8-hour shifts.

Period	1	2	3	4
Number of accidents	138	127	165	170

Test, at the 5% significance level, whether the proportions of all accidents that occur in the four time periods differ. [6]

3 Ten randomly chosen athletes were coached for a 200 m event. For each athlete, the times taken to run 200 m before and after coaching were measured. The sample mean times before and after coaching were 23.43 seconds and 22.84 seconds respectively. For each athlete the difference, d seconds, in the times before and after coaching was calculated and an unbiased estimate of the population variance of d was found to be 0.548. Stating any required assumption, test at the 5% significance level whether the population mean time for the 200 m run decreased after coaching. [7]

4 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{4}{3x^3} & 1 \leqslant x < 2, \\ \frac{1}{12}x & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the upper quartile of X. [4]

(ii) Find the value of a for which $E(X^2) = aE(X)$. [5]

5 Gloria is a market trader who sells jeans. She trades on Mondays, Wednesdays and Fridays. Wishing to investigate whether the volume of trade depends on the day of the week, Gloria analysed a random sample of 150 days' sales and classified them by day and volume (low, medium and high). The results are given in the table below.

		Day		
		Monday	Wednesday	Friday
Volume	Low	15	13	2
	Medium	23	26	23
	High	12	9	27

Gloria asked a statistician to perform a suitable test of independence and, as part of this test, expected frequencies were calculated. These are shown in the table below.

		Day		
		Monday	Wednesday	Friday
Volume	Low	10.00	9.60	10.40
	Medium	24.00	23.04	24.96
	High	16.00	15.36	16.64

(i) Show how the value 23.04 for medium volume on Wednesday has been obtained. [2]

(ii) State, giving a reason, if it is necessary to combine any rows or columns in order to carry out the test. [1]

The value of the test statistic is found to be 21.15, correct to 2 decimal places.

(iii) Stating suitable hypotheses for the test, give its conclusion using a 1% significance level. [4]

Gloria wishes to hold a sale and asks the statistician to advise her on which day to hold it in order to sell as much as possible.

(iv) State the day that the statistician should advise and give a reason for the choice. [2]

6 An anthropologist was studying the inhabitants of two islands, Raloa and Tangi. Part of the study involved the incidence of blood group type A. The blood of 80 randomly chosen inhabitants of Raloa and 85 randomly chosen inhabitants of Tangi was tested. The number of inhabitants with type A blood was 28 for the Raloa sample and 46 for the Tangi sample. The anthropologist calculated 90% confidence intervals for the population proportions of inhabitants with type A blood. They were (0.262, 0.438) for Raloa and (0.452, 0.630) for Tangi, where each figure is correct to 3 decimal places. It is known that 43% of the world's population have type A blood.

(i) State, giving your reasons, whether there is evidence for the following assertions about the proportions of people with type A blood.

(a) The proportion in Raloa is different from the world proportion.

(b) The proportion in Tangi is different from the world proportion.

[3]

(ii) Carry out a suitable test, at the 2% significance level, of whether the proportions of people with type A blood differ on the two islands. [8]

7 A queue of cars has built up at a set of traffic lights which are at red. When the lights turn green, the time for the first car to start to move has a normal distribution with mean 2.2 s and standard deviation 0.75 s. This time is the *reaction time* for the first car. For each subsequent car the reaction time is the time taken for it to start to move after the car in front starts to move. These reaction times have identical normal distributions with mean 1.8 s and standard deviation 0.70 s. It may be assumed that all reaction times are independent.

(i) Calculate the probability that the reaction time for the second car in the queue is less than half of the reaction time for the first car. [6]

(ii) Calculate the probability that the fifth car in the queue starts to move less than 10 seconds after the lights turn green. [5]

(iii) State where, in part **(i)**, independence is required. [1]

8 Two machines, A and B, produce metal rods. Machine B is new and it is required that its accuracy should be checked against that of machine A. The observed variable is the length of a rod. Random samples of rods, 40 from machine A and 50 from machine B, are taken and their lengths, x_A cm and x_B cm, are measured. The results are summarised by

$$\Sigma x_A = 136.48, \quad \Sigma x_B = 176.35, \quad \Sigma x_B^2 = 630.1940.$$

The variance of the length of the rods produced by machine A is known to be $0.0490\,\text{cm}^2$. The mean lengths of the rods produced by the machines are denoted by μ_A cm and μ_B cm respectively.

(i) Test, at the 5% significance level, the hypothesis $\mu_B > \mu_A$. [7]

(ii) Find the set of values of a for which the null hypothesis $\mu_B - \mu_A = 0.025$ would **not** be rejected in favour of the alternative hypothesis $\mu_B - \mu_A > 0.025$ at the a% significance level. [4]

(iii) For the test in part **(i)** to be valid,

(a) state whether it is necessary to assume that the two population variances are equal, [1]

(b) state, giving a reason, whether it is necessary to assume that the lengths of rods are normally distributed. [2]

FINAL MARK SCHEME

STATISTICS 3 4734 June 2006

1	Add two Poisson distributions	M1		
	With mean 17	A1		
	$P(27)=e^{-17}17^{27}/27!$ or $P(\leq 27)-P(\leq 26)$	M1		Use formula or table
	0.00634 or 0.0063, 0.0064 from tables	A1	**4**	M1A1 0.0052 from N(17,17)
2	$H_0: p_1=p_2=p_3=p_4$,			Indication of equality of proportions
	(H_1: They are not all equal)	B1		
	Expected values under H_0=150	B1		
	$X^2= (12^2+23^2+15^2+20^2)/150$	M1		At least one correct term
	=8.653	A1		Accept art 8.65 or 8.66
	Critical value with 3 d.f. = 7.815	B1		
	($X^2 > 7.185$ so) reject H_0 and accept that proportions are different.	B1√	**6**	ft critical value
	SR: For testing one p against another: B0B0M1A1B1(If NH rejected)			
3	Assume population of differences has a normal distribution.			
	or sample random	B1		Either assumption.
	$H_0: \mu_B - \mu_A=0$, $H_1: \mu_B - \mu_A > 0$	B1		AEF
	t=(23.43-22.84)/ √(0.548/10)	M1		
	=2.520	A1		
	CV=1.833	B1		Seen
	2.52 > CV so reject H_o	M1		Allow from CV 2.262 (2-tail), 1.812,1.734
	Accept that there is evidence that mean time has reduced.	A1 √	**7**	ft wrong CV
4 (i)	EITHER: $\int_{q_3}^{4} \frac{1}{12}x\,dx=\frac{1}{4}$ or $\int_{1}^{2} \frac{4}{3x^3}dx+\int_{2}^{q_3} \frac{1}{12}x\,dx=\frac{3}{4}$	M1*		
	$[x^2/24]$ OR $[-2/(3x^2)]+[x^2/24]$	A1		Either
	$(16-q_3^2)/24=1/4$ or $1/3 + q_3^2/ 24 = ¾$	dep *M1		Form equation and attempt to solve
	$q_3=\sqrt{10}$	A1	**4**	Accept to 3 SF
	If they find F(x): M1A1, M1A1			
(ii)	$E(X^2)=\int_1^2 \frac{4}{3x}dx+\int_2^4 \frac{x^3}{12}dx$			
	$E(X)=\int_1^2 \frac{4}{3x^2}dx+\int_2^4 \frac{x^2}{12}dx$	M1		Either correct
	$\left[\frac{4}{3}\ln x\right]_1^2+\left[\frac{x^4}{48}\right]_2^4$	A1		
	$\left[\frac{-4}{3x}\right]_1^2+\left[\frac{x^3}{36}\right]_2^4$	A1		
	a= E(X^2)/E(X)	M1		
	a=2.6659, 2.67	A1	**5**	Or exact value, (3ln2)/5 + 9/4 or equiv.

FINAL MARK SCHEME

STATISTICS 3 4734 June 2006

5	(i)	(48×72/150) or (48/150)(72/150)×150	M1		Multiply and divide relevant values
			A1	**2**	All correct
	(ii)	No, no expected value less than 5	B1	**1**	
	(iii)	H_0:Volume and day are independent			Attributes specified
		(H_1:Volume and day are not independent**)**	B1		
		Critical value for 4 df=13.28	B1		
		Test statistic > 13.28, reject H_0	M1		
		Accept that volume and day are not independent	A1	**4**	
		SR: If B0 in (ii) and classes combined then B1 B1 for 9.21 B1 completion, max 3/4			
	(iv)	Choose Friday	B1		
		Highest volume	B1	**2**	Not reference to E values
6	(i)	(a) No	B1		Must be with reason
		0.43 belongs to relevant interval	B1		
		(b)Yes			
		0.43 is outside relevant interval	B1	**3**	
	(ii)	H_0:$p_R = p_T$, H_1:$p_R \neq p_T$	B1		Proportions
		Estimate of p = 74/165	B1		
		Variance estimate of difference			
		$=(\frac{74}{165})(\frac{91}{165})(\frac{1}{80}+\frac{1}{85})$	B1		May be implied by later work
		z=(28/80-46/85)/σ_{est}	M1		Standardising
			A1		Completely correct expression
		= -2.468	A1		+ or - , 2.47
		Compare correctly with CV			
		-2.468<-2.326, or 2.468 > 2.326	M1		
		Reject H_0 and accept that the proportions differ on the island.	A1	**8**	Conclusion in context
		SR: From variance $p_1q_1/80+ p_2q_2/85$ B1B0B1M1A0A1(2.52)M1A1 Max 6/8			SR: 1-tail test: B0B1B1M1A1A1M0A0 Max
7	(i)	T_1~ N(2.2,0.75^2), T_2 ~ N(1.8,0.70^2)			
		Use T_2 - ½ T_1 normal	M1		Or ½T_1 – T_2
		μ=0.7	A1		
		σ^2=0.7^2+ ¼×0.75^2 (0.630625)	A1		
		(0- μ)/σ	M1		From reasonable σ^2 not just sum
		-0.881	A1		+ or -
		Probability 0.189	A1	**6**	

FINAL MARK SCHEME

STATISTICS 3 4734 June 2006

	(ii)	Use sum of 5 Ts	M1		
		μ=9.4	A1		
		σ^2=2.5225	A1		
		$z=(10-\mu)/\sigma$	M1		Standardising, must be σ
		Probability 0.6473,0.647	A1	**5**	
	(iii)	Calculation of variance	B1	**1**	
8	**(i)**	$s_B^2 = \frac{1}{49}(630.194 - \frac{176.35^2}{50})$	M1		Any equivalent formula
		=0.1675	A1		May be implied by later work
		H_0: $\mu_B - \mu_A = 0$, H_1: $\mu_B - \mu_A > 0$	B1		aef
		z=0.115/√(0.049/40 + 0.1675/50)	M1		Standardising but not from pooled variance estimate
		=1.700	A1		art 1.70
		z > 1.645 , reject H_0	M1		Compare correctly with 1.645
		and accept that $\mu_B > \mu_A$	A1 √	**7**	ft their calculated z
	(ii)	z = 0.09/√(0.004575)	M1		Correct form
		= 1.331	A1		
		H_0 not rejected for $\alpha < 9.16$	M1 A1	**4**	Accept < 9.2, ≤ 9.2 . M1 for correct method for 9.2, A1 for inequality
	(iii)	(a) Not necessary	B1		Ignore any reason
		(b) Not necessary since samples large enough for CLT to be applied (normality of sample means giving normality of difference)	M1		Mention of CLT implied by "sample large"
			A1	**3**	Sample mean (approx) normal. (Do not award if population or sample said to be normal)