RECOGNISING ACHIEVEMENT

ADVANCED GCE **4734/01**

MATHEMATICS

Probability & Statistics 3

MONDAY 16 JUNE 2008 Afternoon

Time: 1 hour 30 minutes

Additional materials (enclosed): None

Additional materials (required):
Answer Booklet (8 pages)
List of Formulae (MF1)

INSTRUCTIONS TO CANDIDATES

- Write your name in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- 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**.
- **You are reminded of the need for clear presentation in your answers.**

This document consists of **4** printed pages.

1 The lengths of rivets produced by a certain factory are checked each day by measuring a random sample of 100 rivets. For a particular day's sample the lengths, x mm, are summarised by $\Sigma x = 761.2$ and $\Sigma x^2 = 6115.04$. The mean and standard deviation of the lengths of all rivets produced that day are denoted by μ mm and σ mm respectively.

(i) Find an unbiased estimate of σ^2. [2]

(ii) Calculate a 95% confidence interval for μ. [3]

(iii) Explain what distributional assumptions (if any) are required for the validity of your calculated confidence interval. [1]

2 The saturated fat content of a particular brand of olive spread is monitored regularly in order to maintain a mean percentage content of 12.6. This is carried out by measuring the saturated fat content in random samples of 10 cartons. For a particular sample, the sample mean and an unbiased estimate of the population variance are calculated. The unbiased estimate of the population variance is 0.1195. It may be assumed that percentage fat content has a normal distribution. Find the critical region for a test at the 10% significance level of whether the population mean percentage fat content exceeds 12.6. [5]

3 A large sample of people were surveyed and classified by 4 levels of income and by which of 3 newspapers they read. The results were arranged in a contingency table consisting of 4 columns and 3 rows. In a χ^2 test of independence between income and choice of newspaper, it was found necessary to combine two of the columns. The value of the test statistic was 12.32.

(i) State a suitable null hypothesis for the test. [1]

(ii) Determine the largest significance level, obtained from tables or calculator, for which independence would be accepted. [3]

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

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

(i) Find $P(X < 2)$. [3]

(ii) Show that the median of X exceeds 1. [3]

(iii) Find $E(X)$. [3]

(iv) Show that $Var(X)$ is not finite. [3]

5 The proportion of syringes of brand A that are faulty is 2.2%. The corresponding proportion for brand B is 2.5%. Random samples of 75 brand A and 90 brand B syringes are taken and the total number of faulty syringes is denoted by X.

(i) Show that the distribution of X can be approximated by a Poisson distribution, and state its mean. [5]

(ii) Find $P(X > 5)$. [2]

6 The proportion of teapots with faulty spouts produced in a factory is denoted by p. In a random sample of 50 teapots, the number with faulty spouts was found to be 6.

(i) Find a 98% confidence interval for p. [4]

(ii) Find an estimate of the sample size for which the sample proportion would differ from p by less than 0.05 with 98% confidence. [5]

7 A psychologist believed that teenage boys worry more than teenage girls and he devised a questionnaire to examine his belief. He gave the questionnaire to a random sample of 24 girls and a random sample of 18 boys. The scores, x_G and x_B for the girls and boys, are summarised by $\Sigma x_G = 1526.8$ and $\Sigma x_B = 1238.4$. Unbiased estimates of the respective population variances, obtained from the samples, are $s_G^2 = 86.79$ and $s_B^2 = 93.01$. Larger scores indicate greater levels of worry.

(i) State two assumptions required for the validity of a t-test to examine the psychologist's belief. [2]

(ii) Comment on one of these assumptions in the light of the data. [1]

(iii) Carry out the test at the 5% significance level. [9]

[Question 8 is printed overleaf.]

8 The numbers of goals scored by my local football team in 80 matches are summarised in the following table.

Number of goals	0	1	2	3	4	5	⩾6
Number of matches	11	15	33	16	2	3	0

(i) Show that the mean of the distribution is 1.9, and find the variance of the distribution. [3]

(ii) Without carrying out a test, explain whether the values of the mean and variance indicate that a Poisson distribution could be a suitable model for the number of goals scored in a match. [2]

The table below gives the expected frequencies, correct to 2 decimal places, for a χ^2 goodness of fit test of a Poisson distribution.

Number of goals	0	1	2	3	4	⩾5
Expected frequency	11.97	22.73	21.60	13.68	6.50	3.52

(iii) Show how the value 13.68 for 3 goals is obtained. [2]

(iv) Stating a required assumption regarding the data, carry out the test at the 5% significance level. Does the outcome of the test confirm your answer to part **(ii)**? [8]

(v) Without further calculation, state two ways in which the test would be different if it were a goodness of fit test of the distribution Po(2), also at the 5% significance level. [2]

4734 Probability & Statistics 3

1 (i)	$\frac{1}{99}(6115.04-\frac{761.2^2}{100})$	M1		AEF
	=3.240	A1	2	
(ii)	761.2/100±z√(3.24/100)	M1		z= 1.282, 1.645, or 1.96
	z = 1.96	B1		
	(7.26,7.96)	A1	3	Allow from σ^2=3.21; allow 7.97 but not from wrong σ. Allow 4 or 5 SF but no more**.**
(iii)	None necessary, since sample size large enough for sample mean to have a normal distribution	B1	1	OR:None necessary, n large enough for Central Limit theorem to apply
			[6]	
2	$(\bar{x}-12.6)/\sqrt{0.1195/10}$	M1		Any variable, correct mean, /10, ignore z
		A1		All correct
	1.383 seen	B1		
	Solve for variable	M1		Allow any symbol (<,>,=)
	$\bar{x}\geq 12.75$	A1	5	Allow > ; 12.7 or 12.8 No z seen
			[5]	
3(i)	Choice of newspaper is independent of level of income	B1	1	Or equivalent
(ii)	Use df=4	B1		May be implied by 13.28 seen or 0.0152
	EITHER: CV 13.28, from df=4 or sig. level	M1		From tables
	Largest significance level is 1%	**B1**		Accept 0.01
	OR: UseP(χ^2 > 12.32)			Use of calculator
	Largest significance level is 1.52%	**B2**	3	Accept 0.0152
			[4]	
		SR: from df=6: CV 12.59 used ; SL=5% : B0M1B1		
4(i)	$\int_0^1 \frac{4}{3}x^3\,dx+\int_1^2 \frac{4}{3x^3}\,dx$ Limits seen anywhere	M1		For both integrals OR 1 - $\int_2^\infty \frac{4}{3x^3}\,dx$
	$\left[\frac{x^4}{3}\right]_0^1+\left[-\frac{2}{3x^2}\right]_1^2$	A1		For both OR $1-\left[-\frac{2}{3x^2}\right]_2^\infty$
	$^5/_6$	A1	3	
(ii)	EITHER: $\int_0^1 \frac{4}{3}x^3\,dx=\frac{1}{3}$	M1		
	< ½	A1		
	Median must exceed 1	A1		
	OR:	M1		Attempt to find median
	m=√(4/3)	A1		M0 for $1.5^{1/4}$
	> 1 AG	A1	3	Accept 1.15..

(iii)	$\int_0^1 \frac{4}{3}x^4\,dx + \int_1^\infty \frac{4}{3x^2}\,dx$	M1		Correct form for at least one integral
	$[4x^5/15] + [-4/(3x)]$	B1		Both integrals correct without limits
	1.6	A1	**3**	AEF
(iv)	$E(X^2) = ... + \int_1^\infty \frac{4}{3x}\,dx$	M1		For second integral
	Second integral = $\left[\frac{4}{3}\ln x\right]_1^\infty$	A1		
	This is not finite, (so variance not finite)	A1	**3** [12]	AEF
5 (i)	Justify a relevant Poisson approximation	M1		Using $n>50$ or n large; $np<5$ or p small (<0.1) or $np \approx npq$
	$E(A)=75\times0.022$ (=1.65), $E(B) = 90\times0.025$ (=2.25)	B1B1		
	Sum of two independent Poisson variables X has a Poisson distribution	A1		
	Mean m =3.9	B1	**5**	Accept Po(3.9)
(ii)	$1 - P(\leq 5)$	M1		Or From Po(m) Accept ≤ 4; OR Exact 1 – sum of at least 5 correct terms
	0.1994	A1	**2** **[7]**	From calculator or tables, art 0.20
6 (i)	Use $p_s \pm zs$	M1		
	z=2.326	B1		
	$s = \sqrt{(0.12\times0.88/50)}$	A1		Or /49
	(0.013,0.227) Allow limits if penalised in Q1	A1	**4**	Or (0.012,0.228) from 49
(ii)	$z(0.12\times0.88/n)^{1/2}$	M1		Any z
	< 0.05	A1		Allow =
	Solve to obtain	M1		Must contain $\sqrt{n}$
	$n > 228.5$	A1		Accept =
	$n \approx 229$ or 230	A1	**5** **[9]**	Must be integer
7 (i)	Each population of test scores should have normal distributions	B1		OR: Variances equal and normal distns B1; Context B1
	with equal variances	B1	**2**	
(ii)	EITHER:Cannot test for normality from data OR: Sample variances are close enough to accept population variances equal	B1	**1**	Not variances are not equal

(iii)	$H_0: \mu_B = \mu_G$, $H_1: \mu_B > \mu_G$	B1	For both. No other variables. Allow words
	$s^2 = (23\times86.79+17\times93.01)/40$	M1	Finding pooled estimate of variance
	$=89.4335$	A1	May be implied by later value of t
	$t = (1238.4/18 - 1526.8/24)/\ [s^2(18^{-1}+24^{-1})]^{1/2}$	M1	With pooled estimate of variance
		A1	All correct
	$= 1.758$	A1	art 1.76, or -
	Use CV of 1.684	B1	Consistent
	$1.758 > 1.684$	M1	Compare correctly with their CV (t value)
	Reject H_0 and accept there is sufficient evidence at the 5% significance level that teenage boys worry more, on average than teenage girls.	A1√ **9**	Not assertive Ft on their 1.758 SR:Using s^2= 93.01/18+86.79/24 : B1M0A0M1A0A1(for 1.749) B1M1(from 1.645 or 1.684)A1 Max 6/9
		12]	
8 (i)	$\Sigma xf/80 = 1.9$ AG	B1	With evidence
	$\Sigma x^2f/80 - 1.9^2$	M1	Or × 80/79
	1.365 or 1.382	A1 **3**	
(ii)	Poisson distribution requires equal mean and variance	B1	May be indicated
	EITHER: No, mean and variance differ significantly OR: Yes, indicated by sample statistics taking into account sampling error	B1 **2**	
(iii)	$e^{-1.9}1.9^3/3!$	B1	Or from tables
	×80	B1 **2**	
(iv)	Considering sample as random selection of all similar matches	B1	
	H_0: Poisson suitable model		
	Combine last two cells	B1	
	$0.97^2/11.97+7.73^2/22.73+11.40^2/21.60$	M1	Any two correct
	$+ 2.32^2/13.68+5.02^2/10.02$	A1	All correct
	=11.63	A1	art 11.6
	CV 7.815	B1 *dep	OR p=0.00875
	$11.63 > 7.815$	M1dep*	OR $0.00875 < 0.05$
	There is sufficient evidence that a Poisson distribution is not a suitable model confirming (or not) the answer to part (ii)	A1√ **8**	Ft (ii) SR: If last cells not combined:$\chi^2 = 12.3$ M1A1A1 CV=9.448 or p = 0.0152, B1*dep the M1dep*
(v)	E-values or probabilities would change	B1	Or other valid observation
	df would increase by 1	B1 **2**	Or CV would change
		[17]	