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.

[Turn over

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

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

(i) Find $\mathrm{P}(X < 2)$. [3]

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

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

(iv) Show that $\mathrm{Var}(X)$ 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 $\mathrm{P}(X > 5)$. [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	$\geqslant 6$
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	$\geqslant 5$
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]