OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS **2643**

Probability & Statistics 3

Friday **23 JANUARY 2004** Morning 1 hour 20 minutes

Additional materials:
Answer booklet
Graph paper
List of Formulae (MF8)

TIME 1 hour 20 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 60.
- 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 A random sample of 50 observations of a random variable X is taken from a population. The data can be summarised by

$$\Sigma x = 1180.0, \qquad \Sigma(x-\bar{x})^2 = 1324.96.$$

Calculate a 99% confidence interval for the population mean μ. [5]

2 X_1 and X_2 are two independent observations from a Poisson distribution with mean 1.6. Find

(i) $\mathrm{P}(X_1 + X_2 \leqslant 9)$, [2]

(ii) $\mathrm{P}(3X_1 < 10)$. [3]

3 A random sample of 60 people who had been treated for drug abuse were monitored after their treatment. These people were classified according to whether they had received more or less than 15 years of full-time education, and the numbers in each category who were subsequently convicted of a criminal offence were also recorded. The results are given in the following table.

	Full-time education	
	15 years or more	Less than 15 years
Subsequent conviction	16	20
No subsequent conviction	6	18

Test, at the 5% significance level, the hypothesis that subsequent conviction is independent of the amount of full-time education received. [7]

4 A commercial grower of tomatoes usually uses a peat-based compost for growing his tomatoes, and he knows that the mean weight of tomatoes produced by a plant is 5.1 kg. One year he decides to grow some of his plants in a coir-based compost to see if it works as well as the peat-based compost. He selects a random sample of 20 of the plants grown in coir-based compost. The weights, x kg, produced by these 20 plants can be summarised by

$$\Sigma x = 92.2, \qquad \Sigma x^2 = 451.36.$$

The weight of tomatoes produced by a plant may be assumed to be normally distributed.

(i) Use a t-test to test, at the 1% significance level, whether the mean weight of tomatoes produced by plants grown in the coir-based compost is less than 5.1 kg. [7]

(ii) Give two reasons why a t-test should be used rather than a test requiring the use of normal distribution tables. [2]

5 A canal boat is moored in a boatyard. When it leaves the yard to go south down the canal it has to pass through a lock and then to travel one mile before it reaches a canal-side pub. The times, in minutes, taken by canal boats to pass through a lock, and to travel one mile without locks, are both normally distributed. The means and standard deviations are as follows.

	Mean	Standard deviation
Time through lock	16	4
Time to travel one mile	19	2

(i) Assuming that the times taken to travel one mile and to pass through a lock are independent of each other, find the probability that it takes the boat less than 40 minutes to get from the boatyard to the pub. [4]

When the boat leaves the pub to go south to the next boatyard it has to pass through three locks and travel for two miles. The first lock is immediately next to the pub, the second lock is reached after travelling one mile and the third after travelling another mile, just before entering the next boatyard. The boat leaves the pub at 2.30 pm.

(ii) Assuming that the times taken to go through the three locks and to travel each of the two miles are all independent of each other, find the probability that the canal boat arrives at the next boatyard between 3.45 and 4 pm. [6]

6 In a village a survey was carried out to estimate the proportion of shoppers within the village who use the small supermarket in the village for their food shopping. The random sample included 97 shoppers under the age of 60, of whom 15 did their food shopping at the supermarket, and 93 shoppers over the age of 60, of whom 27 did their food shopping at the supermarket.

(i) Calculate a 95% confidence interval for the overall proportion of shoppers in the village that do their food shopping at the supermarket in the village. [5]

(ii) Carry out an appropriate statistical test, at the 1% significance level, to test if there is a difference between the proportion of shoppers in the village under 60 who do their food shopping at the village supermarket and the proportion of shoppers over 60 who do their food shopping at the village supermarket. [6]

[Question 7 is printed overleaf.]

7 A continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{4} & -2 \leqslant x < 0, \\ \frac{1}{36}(9 - x^2) & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the cumulative distribution function. [4]

A statistician used this probability density function as a model for the following data observed in an experiment.

Interval	$-2 \leqslant x < -1$	$-1 \leqslant x < 0$	$0 \leqslant x < 1$	$1 \leqslant x < 2$	$2 \leqslant x \leqslant 3$
Observed frequency	17	17	16	8	2

A table of expected frequencies given by the model, correct to 1 decimal place, is given below.

Interval	$-2 \leqslant x < -1$	$-1 \leqslant x < 0$	$0 \leqslant x < 1$	$1 \leqslant x < 2$	$2 \leqslant x \leqslant 3$
Expected frequency	15.0		14.4		

(ii) Copy and complete this table of expected frequencies. [4]

(iii) Test, at the 10% significance level, whether the model is a good fit to the data. [5]

S3 Jan 2004

① Unknown X distribution so use CLT Normal:

C.I. is $\frac{1180}{50} \pm 2.576\sqrt{\frac{1324.96}{50}} \times \sqrt{\frac{50}{49}} \div \sqrt{50}$

$= (21.7, 25.5)$ ⑤

② $P(X_1+X_2 \le 9) = \underline{0.9982}$ ②

$P(X_1 < \frac{10}{3})$. but X_1 is integer valued, so

$= P(X_1 \le 3) = \underline{0.9212}$ ③

③ o_i:

16	20	36
6	18	24
22	38	60

H_0: conviction & time a.ed. are indep.

H_1: " " not

e_i:

13.2	22.8
8.8	15.2

$\chi^2_1 = \sum \frac{(|o_i - e_i| - \frac{1}{2})^2}{e_i} = 1.58$

χ^2_1 crit $= 3.841$. $1.58 << 3.841$

So there's no particular reason to think they are linked from these data. ⑦

④ $H_0: \mu = 5.1$

$H_1: \mu < 5.1$

$t_{19} = \dfrac{\frac{92.2}{20} - 5.1}{\sqrt{\left[\frac{451.36}{20} - \left(\frac{92.2}{20}\right)^2\right] \times \frac{20}{19}} / \sqrt{20}} = -1.86$

$t_{crit} = -2.539$.

retain H_0, i.e. at 1% level there's not compelling evidence to suggest that $\mu < 5.1$. ⑦

Small sample, unknown variance ②

⑤ $L+R \sim N(35, 4^2+2^2)$

$P(L+R < 40) = \Phi\left(\frac{40-35}{\sqrt{20}}\right) = \Phi(1.118\ldots)$

$= \underline{0.868}$ ④

$L_1+L_2+L_3+R_1+R_2 \sim N(86, 56)$

$Pr(3.45 < \text{arrival} < 4pm)$

$= \Phi\left(\frac{90-86}{\sqrt{56}}\right) - \Phi\left(\frac{75-86}{\sqrt{56}}\right)$

$= \Phi(0.534\ldots) - \Phi(-1.4699\ldots) = \underline{0.633}$ ⑥

⑥ 42 out of 190 do.

$P\left(-1.96 < \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} < 1.96\right) = 95\%$. ⑤

So 95% c.i. for p is (~~22.1~~ 16.2%, 28.0%)

$H_0: p_x = p_y$

$H_1: p_x \ne p_y$

$\frac{X}{n_x} - \frac{Y}{n_y}$ is the statistic to use.

May use pooled estimate of common variance as $.00090625\ldots$

$\therefore z = \dfrac{\frac{27}{93} - \frac{15}{97} - 0}{\sqrt{\frac{42}{190} \cdot \frac{148}{190}}\sqrt{\frac{1}{93} + \frac{1}{97}}} = 2.25$ ⑥

So reject H_0 — very clear evidence that over 60's have higher proportion.

⑦ $F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{4}x + \frac{1}{2} & -2 \le x < 0 \\ \frac{1}{4}x + \frac{1}{2} - \frac{1}{108}x^3 & 0 \le x \le 3 \\ 1 & x > 3 \end{cases}$ ④

~~0.25~~	~~0.185~~	~~0.074~~
15	11.1	4.4

combine.

o_i:	17	17	16	10
e_i:	15	15	14.4	15.6

$\therefore \chi^2_3 = 2.72$

$\chi^2_{crit} = 6.251$

So retain H_0, this appears to be a good fit. ⑤