OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 2643

Probability & Statistics 3

Tuesday **28 MAY 2002** Afternoon 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 3 printed pages and 1 blank page.

Registered Charity Number: 1066969

1 A car repair firm receives call-outs both as a result of breakdowns and also as a result of accidents. On weekdays (Monday to Friday), call-outs resulting from breakdowns occur at random, at an average rate of 6 per 5-day week; call-outs resulting from accidents occur at random, at an average rate of 2 per 5-day week. The two types of call-out occur independently of each other.

Find the probability that the total number of call-outs received by the firm on one randomly chosen weekday is more than 3. [5]

2 A random sample of 80 precision-engineered cylindrical components is checked as part of a quality control process. The diameters of the cylinders should be 25.00 cm. Accurate measurements of the diameters, x cm, for the sample are summarised by

$$\Sigma(x - 25) = 0.44, \qquad \Sigma(x - 25)^2 = 0.2287.$$

Calculate a 99% confidence interval for the population mean diameter of the components. [6]

3 Boxes of matches contain 50 matches. Full boxes have mean mass 20.0 grams and standard deviation 0.4 grams. Empty boxes have mean mass 12.5 grams and standard deviation 0.2 grams. Stating any assumptions that you need to make, calculate the mean and standard deviation of the mass of a match. [7]

4 The lengths of time, in seconds, between vehicles passing a fixed observation point on a road were recorded at a time when traffic was flowing freely. The frequency distribution in Table 1 is a summary of the data from 100 observations.

Time interval (x seconds)	$0 < x \leqslant 5$	$5 < x \leqslant 10$	$10 < x \leqslant 20$	$20 < x \leqslant 40$	$40 < x$
Observed frequency	49	22	20	7	2

Table 1

It is thought that the distribution of times might be modelled by the continuous random variable X with probability density function given by

$$f(x) = \begin{cases} 0.1e^{-0.1x} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Using this model, the expected frequencies (correct to 2 decimal places) for the given time intervals are shown in Table 2.

Time interval (x seconds)	$0 < x \leqslant 5$	$5 < x \leqslant 10$	$10 < x \leqslant 20$	$20 < x \leqslant 40$	$40 < x$
Expected frequency	39.35	23.87	23.25	11.70	1.83

Table 2

(i) Show how the expected frequency of 23.87, corresponding to the interval $5 < x \leqslant 10$, is obtained. [4]

(ii) Test, at the 10% significance level, the goodness of fit of the model to the data. [5]

5 The continuous random variable X has (cumulative) distribution function given by

$$F(x) = \begin{cases} 1 - \dfrac{1}{(x+1)^3} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the probability density function of X. [2]

(ii) Show that $E(X + 1) = \frac{3}{2}$, and hence write down the value of $E(X)$. [4]

(iii) Find $\text{Var}(X + 1)$, and hence write down the value of $\text{Var}(X)$. [3]

6 A factory manager wished to compare two methods of assembling a new component, to determine which method could be carried out more quickly, on average, by the workforce. A random sample of 12 workers was taken, and each worker tried out each of the methods of assembly. The times taken, in seconds, are shown in the table.

Worker	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Time in seconds for Method 1	48	38	47	59	62	41	50	52	58	54	49	60
Time in seconds for Method 2	47	40	38	55	57	42	42	40	62	47	47	51

(i) (a) Carry out an appropriate t-test, using a 2% significance level, to test whether there is any difference in the times for the two methods of assembly. [7]

(b) State an assumption needed in carrying out this test. [1]

(ii) Instead of using the same 12 workers to try both methods, the factory manager could have used two independent random samples of workers, allocating Method 1 to the members of one sample and Method 2 to the members of the other sample.

(a) State one disadvantage of a procedure based on two independent random samples. [1]

(b) State any assumptions that would need to be made to carry out a t-test based on two independent random samples. [2]

7 Certain types of food are now sold in metric units. A random sample of 1000 shoppers was asked whether they were in favour of the change to metric units or not. The results, classified according to age, were as shown in the table.

	Age of shopper Under 35	35 and over	Total
In favour of change	187	161	348
Not in favour of change	283	369	652
Total	470	530	1000

(i) Use a χ^2 test to show that there is very strong evidence that shoppers' views about changing to metric units are not independent of their ages. [6]

(ii) The data may also be regarded as consisting of two random samples of shoppers; one sample consists of 470 shoppers aged under 35, of whom 187 were in favour of change, and the second sample consists of 530 shoppers aged 35 or over, of whom 161 were in favour of change. Determine whether a test for equality of population proportions supports the conclusion in part (i). [7]

S3 jun 02

1. No of callouts per day for breakdown $\sim P(1.2)$
No of callouts per day for accident $\sim P(0.4)$
Using sum of Poissons=Poisson,
Total no. of callouts per day $\sim P(1.6)$

$P(\text{total no}> 3) = 1-(e^{-1.6}+1.6e^{-1.6}+\frac{1.6^2e^{-1.6}}{2!}+\frac{1.6^3e^{-1.6}}{3!})$

$= 0.0788$

---

2. Using false origin of 25,
$s^2 = \frac{80}{79}\times\left(\frac{0.2287}{80}-\left(\frac{0.44}{80}\right)^2\right) = 0.0028643\ldots$
n is so large that Normal may be used to model distribution of $\bar{X}$
$P\left(-2.576 < \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} < 2.576\right) \approx 99\%$
which can be rearranged to give the 99% confidence interval for μ as (-0.00991,0.02091), but this is referred to the origin of 25, so the c.i. is $(24.990, 25.021)$

---

3. X = mass of a match in grams.
Y = mass of a matchbox in grams.
Let the mean and standard deviation of X be μ and σ
We are given that
$E\left(\sum_1^{20} X_i + Y\right) = 20$
$V\left(\sum_1^{20} X_i + Y\right) = 0.4^2$
thus
$20\mu + 12.5 = 20$
so $\mu = 0.375$ grams

and $20\sigma^2 + 0.2^2 = 0.4^2$, noting here that all the X's and the Y need to be independent,
so $\sigma = 0.0775$ grams

---

4. (i)
By $100 \times \int_5^{10} 0.1e^{-0.1x}\,dx$
$= -100[e^{-0.1x}]_5^{10}$
$= 100(e^{-0.5}-e^{-1})$
$= 23.87$

(ii) $X^2 = \frac{9.65^2}{39.35} + \frac{1.87^2}{23.87} + \frac{3.25^2}{23.25} + \frac{4.53^2}{13.53}$
(note last two cells needed to be combined because $e_5 < 5$)
$= 4.48$

$\nu = 3$, and $\chi^2_{crit} = 6.251$.

so the data is rather inconclusive: there is not enough evidence to throw out the proposed model, but neither is it a particularly close fit.

---

5. (i) $f(x) = \frac{3}{(x+1)^4} \quad x \geq 0$
$\quad 0 \quad$ otherwise

(ii) Defining Y as $X + 1$, wecan see that

$$E(Y) = \int_1^\infty y \times \frac{3}{y^4} dy$$
$$= \left[-\frac{3}{2y^2}\right]_1^\infty$$
$$= \frac{3}{2}$$

Since $E(X + 1) = E(X) + 1$,
$$E(X) = \frac{1}{2}$$

(iii)
$$V(X + 1) = \int_1^\infty y^2 \times \frac{3}{y^4} dy - \left(\frac{3}{2}\right)^2$$
$$= \left[-\frac{3}{y}\right]_1^\infty - \frac{9}{4}$$
$$= \frac{3}{4}$$

and so $V(X) = \frac{3}{4}$

---

6. (i) A paired-sample t-test is appropriate.
(a) Designating D as the difference between times for one worker,
$H_0: \mu_D = 0$
$H_1: \mu_D \neq 0$

d_i	1	-2	9	4	5	-1	8	12	-4	7	2	9

$$s^2 = \frac{486}{12} - \left(\frac{50}{12}\right)^2 = 23.13888..$$

so

$$t = \frac{4.16666.. - 0}{\sqrt{\frac{23.16666}{12}}}$$
$$= 3.00$$

$\nu = 11$, and $t_{crit} = 2.718$, so there is significant evidence of a difference in the mean times.

(b) The assumption was that differences d_i are distributed Normally.

(ii) It may avoid a systematic trend like the workers becoming tired and taking longer to perform the second task
Assumptions would have to be times distributed Normally (assuming samples still small), with equal variance.

---

7. (i) H_0: the attitude to change and age are independent
H_1: they're not

The e_i's are:

163.56	184.44
306.44	345.56

$$X^2 = \frac{22.94^2}{163.56} + \frac{22.94^2}{184.44} + \frac{22.94^2}{306.44} + \frac{22.94^2}{345.56} = 9.31$$

$\nu = 1$, which would be significant even at the $\frac{1}{2}$% level (χ^2 for 0.995 is 7.879)
i.e. very strong evidence that they are not independent.

(ii) H_0: $p_x - p_y = 0$
H_1: $p_x - p_y \neq 0$

The statistic

$$\frac{p_{sx} - p_{sy} - 0}{\hat{\sigma}\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

is approximately Normal (0,1), where $\hat{\sigma} = \frac{x+y}{n_x+n_y}$.

so

$$z = \frac{0.39787.. - 0.30377.. - 0}{\sqrt{0.348\left(\frac{1}{470} + \frac{1}{530}\right)}} = 2.518$$

Critical value for the 2% level is 2.326, so this would be significant at the 2% level, though not at the 1% level. It's therefore still compelling evidence of a difference in attitude between the age groups.
