OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 4734

Probability & Statistics 3

Specimen Paper

Additional materials:
 Answer booklet
 Graph paper
 List of Formulae (MF 1)

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 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 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.

© OCR 2004 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 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]

3 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

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

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

(ii) For the calculation in part **(i)** to be valid, is it necessary to assume that component diameters are normally distributed? Justify your answer. [2]

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. [5]

(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 a triangular distribution with probability density function given by

$$f(x) = \begin{cases} 1+x & -1 \leqslant x \leqslant 0, \\ 1-x & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that, for $0 \leqslant a \leqslant 1$,

$$P(|X| \leqslant a) = 2a - a^2. \qquad [3]$$

The random variable Y is given by $Y = X^2$.

(ii) Express $P(Y \leqslant y)$ in terms of y, for $0 \leqslant y \leqslant 1$, and hence show that the probability density function of Y is given by

$$g(y) = \frac{1}{\sqrt{y}} - 1, \quad \text{for } 0 < y \leqslant 1. \qquad [4]$$

(iii) Use the probability density function of Y to find $E(Y)$, and show how the value of $E(Y)$ may also be obtained directly using the probability density function of X. [4]

(iv) Find $E(\sqrt{Y})$. [2]

6 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	Age of shopper: 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. [7]

(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]

7 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. [8]

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

(c) Calculate a 95% confidence interval for the population mean time difference for the two methods of assembly. [3]

(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]

1	Model for call-outs is Poisson Mean is $\frac{1}{5}(6+2)$ $=1.6$ Probability is $1-0.9212$ $=0.0788$	B1 M1 A1 M1 A1 **5** $\boxed{5}$	For any implication of Poisson For summing two relevant parameters For correct mean of 1.6 For relevant use of tables For correct answer
2	Assume $F = E + M_1 + M_2 + \ldots + M_{50}$, where the masses of the 50 matches in a box are independent the mass of the empty box is independent of the masses of the matches $20.0 = 12.5 + 50\mu$ Hence mean mass of a match is 0.15 grams $0.4^2 = 0.2^2 + 50\sigma^2$ Hence standard deviation is 0.049 grams	 B1 B1 M1 A1 M1 A1 A1 **7** $\boxed{7}$	(The relation itself may be implied) For one relevant valid assumption For another relevant valid assumption For attempting $\mathrm{E}(F)$ in terms of μ For correct value 0.15 For attempting $\mathrm{Var}(F)$ as a sum For correct equation For correct value 0.049
3	**(i)** $\bar{x} = 25.0055$ $s^2 = \frac{1}{79}\left(0.2287 - \frac{0.44^2}{80}\right)$ $= 0.00286...$ Interval is $25.0055 \pm 2.576\sqrt{\frac{0.00286}{80}}$ Hence $24.99(0) < \mu < 25.02(1)$	B1 M1 A1 M1 B1 A1 **6**	For correct sample mean, or equivalent; the 25 may be taken into account later For correct unsimplified expression For correct unbiased estimate For calculation of the form $\bar{x} \pm z\sqrt{(s^2/n)}$ For relevant use of $z = 2.576$ For correct interval, stated to an appropriate degree of accuracy
	(ii) The sample size of 80 is sufficient large for the Central Limit Theorem to apply, so it is not necessary to assume a normal distribution	M1 A1 **2** $\boxed{8}$	For mention of sample size and CLT For the correct conclusion and reason
4	**(i)** $f_e = 100 \times \int_5^{10} 0.1\mathrm{e}^{-0.1x}\,\mathrm{d}x$ $= 100\left[-\mathrm{e}^{-0.1x}\right]_5^{10}$ $= 100(\mathrm{e}^{-0.5} - \mathrm{e}^{-1}) = 23.87$	M1 A1 M1 M1 A1 **5**	For attempting to integrate $\mathrm{f}(x)$ For correct indefinite integral For multiplying by total frequency For use of correct limits For obtaining given answer correctly
	(ii) Combining: $\begin{array}{ccccc} f_o & 49 & 22 & 20 & 9 \\ f_e & 39.35 & 23.87 & 23.25 & 13.53 \end{array}$ Test statistic is $\frac{9.65^2}{39.35} + \frac{1.87^2}{23.87} + \frac{3.25^2}{23.25} + \frac{4.53^2}{13.53}$ $= 4.484$ This is less than 6.251 Hence there is a satisfactory fit	M1 M1 A1 M1 A1√ **5** $\boxed{10}$	For combining the last two classes For correct calculation process For correct value 4.48 For comparison with the correct critical value For correct conclusion, in terms of the fit

5 **(i)** $P(\lvert X\rvert < a) = P(-a < X < a)$	M1	For consideration of two areas, or equiv
$= \int_{-a}^{0} (1+x)\,dx + \int_{0}^{a} (1-x)\,dx$	A1	For integrals or equivalent trapezia
$= \left[x + \frac{1}{2}x^2\right]_{-a}^{0} + \left[x - \frac{1}{2}x^2\right]_{0}^{a} = 2a - a^2$	A1 **3**	For showing the given answer correctly
(ii) $P(Y \leqslant y) = P(X^2 \leqslant y) = P\left(\lvert X\rvert \leqslant \sqrt{y}\right) = 2\sqrt{y} - y$	M1	For expression of $P\left(X^2 \leqslant y\right)$ in terms of y
	A1	For correct expression $2\sqrt{y} - y$
Hence the pgf of Y is $\frac{d}{dy}\left(2\sqrt{y} - y\right) = \frac{1}{\sqrt{y}} - 1$	M1	For differentiation of previous expression
	A1 **4**	For showing the given answer correctly
(iii) $E(Y) = \int_0^1 y^{\frac{1}{2}} - y\,dy = \left[\frac{2}{3}y^{\frac{3}{2}} - \frac{1}{2}y^{\frac{1}{2}}\right]_0^1 = \frac{1}{6}$	M1	For the correct integral in terms of y
	A1	For correct answer $\frac{1}{6}$
$E(X^2) = \int_{-1}^{0} (x^2 + x^3)\,dx + \int_0^1 (x^2 - x^3)\,dx$	M1	For the correct integrals in terms of x
$= \left[\frac{1}{3}x^3 + \frac{1}{4}x^4\right]_{-1}^{0} + \left[\frac{1}{3}x^3 - \frac{1}{4}x^4\right]_0^1 = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$	A1 **4**	For the correct answer correctly obtained
(iv) $E(\sqrt{Y}) = \int_0^1 y^{\frac{1}{2}}\ g(y)\,dy = \int_0^1 (1 - y^{\frac{1}{2}})\,dy$	M1	For forming the correct integral
$= \left[y - \frac{2}{3}y^{\frac{3}{2}}\right]_0^1 = \frac{1}{3}$	A1 **2**	For the correct answer $\frac{1}{3}$
	$\boxed{13}$	
6 **(i)** H_0: shoppers' views and age are independent,		
H_1: shoppers' views and age are not independent	B1	For stating both hypotheses
Exp frequencies under H_0 are 163.56 184.44 / 306.44 345.56	M1	For correct method for expected frequencies
	A1	For all four correct
Test statistic is $\frac{22.94^2}{163.56} + \frac{22.94^2}{184.44} + \frac{22.94^2}{306.44} + \frac{22.94^2}{345.56}$	M1	For correct calculation process, inc Yates
$= 9.31\ldots$	A1	For correct value of the test statistic
This is greater than the critical 0.5% value of 7.879	M1	For a relevant (1 df) comparison
Hence there is very strong evidence to reject H_0 and conclude that views about changing to metric units are not independent of age	A1√ **7**	For correctly justifying the given answer (the final two marks remain available if Yates' correction is omitted)
(ii) $H_0: p_1 = p_2$, $H_1: p_1 \neq p_2$	B1	For both hypotheses stated
Under H_0 the sample value of the common proportion is $\frac{187+161}{1000} = 0.348$	B1	For correct value of estimated p
Test statistic is $\dfrac{\frac{187}{470} - \frac{161}{530}}{\sqrt{0.348 \times 0.652 \times \left(\frac{1}{470} + \frac{1}{530}\right)}}$	M1	For num $p_1 - p_2$ and denom using attempted s.d. based on a common estimate of p
	A1	For completely correct expression
$= 3.118$	A1	For correct value of the test statistic
This is greater than the 0.2% (two-tail) critical value of 3.090	M1	For a relevant comparison using the normal distribution
Hence this test supports the conclusion of part **(i)**	A1√ **7**	For any relevant comparison or comment
	$\boxed{14}$	

7	(i)	(a)	$H_0: \mu_d = 0,\ H_1: \mu_d \neq 0$	B1		For both hypotheses stated
			$\bar{d} = 4.1667$	B1		For correct mean difference (subtraction can be either way round)
			$s^2 = \frac{486}{11} - \frac{50^2}{11 \times 12} = 25.2424$	M1		For calculation of unbiased variance estimate
				A1		For correct value 25.24...
			Test statistic is $\frac{4.1667 - 0}{\sqrt{(25.2424/12)}}$	M1		For correct standardising process
			$= 2.873$	A1		For correct value of test statistic
			This is greater than the critical value 2.718	M1		For a relevant comparison using t tables
			Hence there is enough evidence to reject H_0 and conclude that there is a difference between the times for the two methods	A1√	**8**	For correctly stated conclusion in context
		(b)	Population of differences is normal	B1	**1**	For correct statement
		(c)	Interval is $4.1667 \pm 2.201\sqrt{\frac{25.2424}{12}}$	M1		For calculation of the form $\bar{d} \pm t\sqrt{(s^2/n)}$
				B1		For relevant use of $t = 2.201$
			Hence $0.97 < \mu_d < 7.36$	A1	**3**	For correct interval
	(ii)	(a)	Variation in the speed of individual workers is not eliminated, and may be large compared with the difference between the methods that is being tested	B1	**1**	For any relevant valid statement
		(b)	Both samples are from normal populations	B1		For a correct statement about normality
			The population variances are equal	B1	**2**	For a correct statement about the variances
					$\boxed{15}$	