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 **11 JUNE 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.

1 At a certain time of day, the number of vehicles approaching a level crossing from the north has a Poisson distribution with mean 1.2 per minute. The number approaching the same level crossing from the south has a Poisson distribution with mean 0.8 per minute. At that time of day the level crossing gates close for 4 minutes to allow a train to pass. Find the probability that when the gates are about to open there are more than 10 vehicles queueing at the gates. [4]

2 The random variable X has mean 12 and variance 9. The random variable Y has mean 7 and variance 2.5.

(i) Assuming that X and Y are independent of each other, find

(a) $\mathrm{E}(X - Y)$,

(b) $\mathrm{Var}(X - Y)$,

(c) $\mathrm{E}(3X + 2Y)$,

(d) $\mathrm{Var}(3X + 2Y)$.

[4]

(ii) State which of the calculations in part **(i)** would not be justified without the assumption of independence. [1]

3 A researcher took a random sample from a population which was normally distributed with known variance σ^2. She calculated a (symmetric) 90% confidence interval for the population mean μ to be $7.5 < \mu < 10.7$.

(i) Calculate the sample mean $\bar{x}$. [1]

She then used the same data to calculate a 95% confidence interval which she found to be $8.5 < \mu < 10.9$.

(ii) Given that her 90% confidence interval was correctly calculated, give two reasons why she could tell that this 95% confidence interval was not correct. [2]

(iii) The original sample size was 25. Find the smallest sample size required to give a 99% confidence interval with total width less than 3 units. [4]

4 A random sample of 200 teachers in Higher Education, Secondary Schools and Primary Schools gave the following numbers of men and women in each sector.

	Higher Education	Secondary Schools	Primary Schools
Men	21	39	20
Women	13	55	52

In a χ^2 test of whether there is an association between age group taught and gender the following expected frequencies were obtained.

	Higher Education	Secondary Schools	Primary Schools
Men	13.6	37.6	28.8
Women	20.4	56.4	43.2

(i) Show how the expected frequency of 56.4 was obtained. [2]

(ii) Calculate the value of $\frac{(O-E)^2}{E}$ for each cell. [2]

(iii) Complete the χ^2 test, and show that there is sufficient evidence to reject, at the 1% significance level, the null hypothesis that age group taught is independent of gender of the teacher. [3]

(iv) Identify which cell makes the greatest contribution to the total value of the test statistic. Hence comment on the association between age group taught and gender. [2]

5 A teacher gives her class of 30 pupils a short multiple-choice test every Friday. The test consists of 6 questions each with a choice of 4 possible answers, only one of which is correct. The results of the class on one particular Friday are summarised in the frequency table given below.

Number of correct answers	0	1	2	3	4	5	6
Number of students	2	7	10	7	2	1	1

(i) Test, at the 5% significance level, the goodness of fit of the model $B(6, \frac{1}{4})$. [9]

(ii) Comment on the suggestion that all the students were guessing. [1]

[Questions 6 and 7 are printed overleaf.]

[Turn over

6 A random variable X has a continuous uniform distribution on the interval $0 < x < \frac{1}{2}\pi$.

(i) Write down an expression for the probability density function of X. [1]

(ii) Find the cumulative distribution function of X. [2]

A beach runs in a straight line from east to west. A boat travels 300 metres from the beach in a straight line at an angle of X radians to the beach, where X is the random variable defined above. The random variable Y is defined by $Y = 300 \sin X$, so that Y represents the distance of the boat from the nearest point on the beach.

(iii) Find the expected distance of the boat from the nearest point on the beach. [4]

(iv) Find an expression for the cumulative distribution function of Y. [3]

(v) Hence, or otherwise, find the probability that the boat is more than 100 metres from the nearest point on the beach. [2]

7 A special diet was recommended to help people lose weight. The losses in weight, d kg, of a random sample of 7 people who followed the recommended special diet are summarised by $\Sigma d = 8.6$, $\Sigma d^2 = 30.28$.

(i) Find a 90% confidence interval for the mean loss of weight of people following this recommended diet, assuming that this sample is taken from a population with a normal distribution. [4]

A slimming clinic ran a series of classes to help people lose weight. The losses in weight, w kg, of a random sample of 9 people who attended this series of classes are summarised by the sample mean, $\bar{w} = 3.17$, and the unbiased estimate of the population variance, $s_w^2 = 4.107$.

(ii) Assuming that this second sample is also taken from a population with a normal distribution and that these samples are independent of each other, test, at the 10% significance level, whether this series of classes produces a greater mean loss of weight than the diet. [7]

(iii) State one other assumption that must be made for this test to be valid and say whether this assumption seems reasonable. [2]

1.			.
Mean = 4(1.2 + 0.8)	M1		Summing 2 relevant parameters
X ~ Po(8)	A1		Poisson seen or used later
P(X > 10) = 1 – 0.8159	M1		Relevant use of tables and 1- P(X $\le$ 9 or 10)
= 0.1841	A1	4	c.a.o.
2. (i)			
(a) 5	B1		c.a.o. in each case.
(b) 11.5	B1		
(c) 50	B1		
(d) 91	B1	4	
(ii) in parts (b) and (d) , independence is required.	B1	1	Both parts required and no others.
3.			
(i) $\bar{x}$ = 9.1	B1	1	c.a.o.
(ii) Would give $\bar{x}$ = 9.7	B1		Wrong value for $\bar{x}$ or equivalent.
Width should be greater than a 90% interval, in fact it is 2.4 compared with 3.2	B1	2	Confidence interval too narrow or lower limit (8.5) should be lower than 7.5
(iii) Interval width for 90% => $1.645.\dfrac{\sigma}{\sqrt{25}}$ =1.6	B1		Equation for σ from 90% interval
99% width => $2.576\,.\dfrac{\sigma}{\sqrt{n}} < 1.5$	B1		Inequality or equation involving 3 or 1.5
=> $\dfrac{\sqrt{n}}{\sqrt{25}} > \dfrac{2.576}{1.645}\cdot\dfrac{1.6}{1.5}$	M1		Eliminating σ from two equations
=> n > 69.7 => n = 70	A1	4	Allow n $\ge$ 70

4.		
(i) $\frac{120x94}{200} = 56.4$ A.G.	M1 A1 2	(Row total x Column total)/Overall total
(ii) $\frac{(O-E)^2}{E} = \frac{(21-13.6)^2}{13.6}$, ... = 4.026.. 0.052.. 2.688.. 2.684.. 0.034.. 1.792..	M1* A1 2	Two correct calculations seen or implied later All correct to 2 dp – allow truncating – allow correct fractions.
(iii) $\sum \frac{(O-E)^2}{E} = 11.3$ 11.3 is greater than the critical χ^2 value of 9.210 Sufficient evidence to reject the null hypothesis that gender is independent of sector taught in.	B1 M1dep* A1 3	Correct to 1 dp. For comparison of $\sum \frac{(O-E)^2}{E}$ value with 9.210 Correct conclusion in context.
(iv) Men in Higher Education . There are more men, and hence less women, teaching in Higher Education than would be expected if the variables were independent.	B1 B1 2	Cell correctly identified Any sensible comment that follows from the identification of the cell.
5. x | 0 1 2 3 ≥ 4 P(X=x) | 0.1780.. 0.3559.. 0.2967.. 0.1318.. 0.0376.. E values | 5.3 10.7 8.9 4.0 1.1 Combining , x ≥ 3, O = 11, E = 5.1 $\sum \frac{(O-E)^2}{E} = \frac{(2-5.3)^2}{5.3} + ...$ = 10.3 or 10.4	M1 A1 A1 M1 A1 M1* A1	Evaluating binomial probabilities for 0 - 3 Probabilities for 0 – 3 correct to 2dp E values for 0 – 3 correct to 1 dp Combining to E ≥ 5 Correct to 1dp At least one correct calculation seen.
10.3 is greater than the critical χ^2 value of 7.815 Reject H_o - the model is not a good fit.	M1dep* A1 9	Comparison of $\sum \frac{(O-E)^2}{E}$ value with 7.815 or consistent with no. of cells. Correct conclusion about the model, in context.
The numbers scoring 2 correct and 3 or more correct is greater than predicted by the model. This suggest that the probability of getting a question correct is better than ¼ - better than guessing.	B1 1	Use of test and data to comment on suggestion that students were guessing .

6.		.
(i) $f(x) = \frac{2}{\pi}$, $(0 < x < \frac{1}{2}\pi)$	B1 1	Accept $1 \div \frac{\pi}{2}$
(ii) $F(x) = \frac{2x}{\pi}$, $0 < x < \frac{1}{2}\pi$	B1 /	/ for kx, for any k - both required.
$F(x) = 0$ for $x < 0$; $F(x) = 1$ for $x > \frac{1}{2}\pi$	B1 2	
(iii) $E(Y) = E((300)\sin X) = (300) \int_0^{\frac{\pi}{2}} \frac{2}{\pi} \sin x \, dx$	M1,	$\int f(x) \sin x \, dx$ - their f(x),
	A1 /	Correct expression – limits now required.
$= \frac{600}{\pi} [- \cos x]$	A1 /	Limits not required
$= \frac{600}{\pi}$ metres (191 m)	A1 4	Exact or correct to 3sf.
(iv) $G(y) = P(300 \sin X < y)$	M1	Expressing G(y) in terms of X, allow ≤
$= P(X < \sin^{-1} \frac{y}{300})$	M1	Converting into $\sin^{-1}$ form
$= \frac{2}{\pi} . \sin^{-1} \frac{y}{300}$ $(0 < y < 300)$	A1 3	c.a.o.
(v) $P(Y > 100) = 1 - \frac{2}{\pi} . \sin^{-1} \frac{100}{300}$	M1	Use of 1 - G(100)
$= 0.784$ (3sf)	A1 2	allow 0.783

7. (i) $S^2 = (\frac{30.28}{7} - (\frac{8.6}{7})^2) . \frac{7}{6} = 3.285$	B1	Accept S = 1.812.. or biased versions 2.815...or 1.678..
90% interval is $\frac{8.6}{7} \pm t.\sqrt{\frac{3.285..}{7}}$	M1	Calculation of form $\bar{x} \pm t\frac{s}{\sqrt{n}}$
with t = 1.943	B1	Relevant use of t value
$-0.10 < \mu < 2.56$ (2 dp)	A1 4	Correct to 2 dp
(ii) $H_0: \mu_w = \mu_d$, $H_1: \mu_w > \mu_d$	B1	
Pooled estimate = $\frac{6(3.285)+8(4.107)}{14}$ (=3.755)	M1	Use of correct formula
$t = \pm \frac{3.17 - \frac{8.6}{7}}{\sqrt{3.754(\frac{1}{7}+\frac{1}{9})}}$	M1*	Correct form of two sample statistic
	A1	Substitution of correct values
$= \pm$ 1.98 or 1.99	A1	
| t | is greater than the critical t value 1.345	M1dep*	Comparison with correct critical value
Sufficient evidence to reject H_0 and conclude loss of weight greater due to classes.	A1 7	Correct conclusion in context
(iii) Population variances equal	B1	Comparison of variances
Sample variances close enough to justify assumption	B1 2	Comment on sample variances
---	---------	--
SR For use of large sample test in (ii)		
$H_0: \mu_w = \mu_d$, $H_1: \mu_w > \mu_d$	B1	
NO pooled estimate	M0	
$z = \pm \frac{3.17 - \frac{8.6}{7}}{\sqrt{\frac{3.285}{7}+\frac{4.107}{9}}}$	M1	Use of formula
	A1	Completely correct substitution
= 2.017...	A0	
|z| > critical z value 1.282	M1	Use of t value not allowed here
Sufficient evidence to reject H_0 and conclude loss of weight greater due to classes.	A1	
	5 / 7	