OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 2643

Probability & Statistics 3

Thursday **5 JUNE 2003** 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 variable X has a Poisson distribution with mean 1.5. A random variable Y is independent of X and has a Poisson distribution with mean 2.6. Find $P(X + Y \geqslant 6)$. [3]

2 Gamblers have been known to use a special six-sided die known as a 'double five'. A fair six-sided die is converted into a 'double five' by changing the face which normally has 2 spots on it by adding three more spots so that now two of the faces show 5 spots. The other four faces still show 1, 3, 4 and 6 spots respectively. Some gamblers believe that the addition of the three spots can affect the balance of the die. In a game a 'double five' was rolled 45 times and the frequencies of the scores obtained on the die were as shown in the following table.

Score on die	1	3	4	5	6
Frequency	10	7	5	18	5

Test, at the 10% significance level, whether the addition of 3 extra spots has affected the balance of the die. [6]

3 A researcher asked a random sample of 11 sixteen-year-olds to draw a straight line of length 30 cm without a ruler. The lengths, x cm, of the lines that they drew were as follows.

25.4 35.7 35.4 33.6 33.1 31.9 30.8 30.2 28.5 28.5 27.9

$[\bar{x} = 31.0, \quad \Sigma(x - \bar{x})^2 = 107.58.]$

The researcher then asked a random sample of 8 people aged between 40 and 50 to draw a straight line of length 30 cm without a ruler. The lengths, y cm, of the lines that they drew were as follows.

39.8 39.6 37.3 35.3 34.4 32.0 29.9 33.3

$[\bar{y} = 35.2, \quad \Sigma(y - \bar{y})^2 = 87.52.]$

It is intended to test, at the 5% significance level, whether there is a difference between the average lengths of lines drawn by people in the two different age groups.

(i) State any assumptions that are required for the test. [2]

(ii) Carry out the test. [7]

4 In 1991 the Government introduced the Council Tax which was then a new method of taxing people on the value of the property that they occupied. In one village, each residential property was classified into one of 4 bands, labelled C, D, E or F, according to the value of the property, band F containing the most valuable properties. In that village, a resident conducted a survey of a random sample of 200 houses. He asked the residents which band their house had been put in. He also recorded other information from a visual inspection of the houses, including whether they had a double garage, a single garage or no garage at all. The results are shown in the table below.

Council Tax Band	C	D	E	F	Total
Double garage	13	19	20	8	60
Single or no garage	55	54	24	7	140
Total	68	73	44	15	200

Test, at the 10% significance level, whether there is an association between the Council Tax Band and the number of garages. [8]

5 A machine is designed to fill bottles with 50 ml of perfume. The machine is serviced at the beginning of each month. Before a monthly service the volume, in millilitres, of perfume put into the bottles was normally distributed with mean 49.8 and variance 0.19. After the service a random sample of bottles was taken and the volume of perfume in each bottle is given below.

Bottle number	1	2	3	4	5	6	7	8	9	10	11	12
Volume (x ml)	49.9	50.1	51.2	49.7	50.1	49.8	49.5	50.2	50.3	50.5	50.2	50.9

[$\Sigma x = 602.4$, $\Sigma x^2 = 30\,243.08$.]

After the service the volume of perfume in a bottle has the distribution $N(\mu, \sigma^2)$. Calculate a 98% confidence interval for μ, giving your answer correct to 2 decimal places,

(i) assuming that $\sigma^2 = 0.19$, [4]

(ii) using the data above to estimate the value of σ^2. [4]

A two-tailed test is carried out, at the 2% significance level, of the null hypothesis that, after the service, the mean volume of perfume in a bottle is 50.0 ml, assuming that the population variance is unknown.

(iii) Without further calculation, state, with a reason, what the conclusion of that hypothesis test would be. [2]

6 In a supermarket there is an 'Express' checkout for customers with no more than five items and the supermarket promises that customers will wait no longer than 5 minutes before they leave the checkout. The time, in seconds, spent at the checkout by a customer may be assumed to be normally distributed with mean 33 and standard deviation 5. Sally joins a queue in which there are 7 other customers, of whom the first is just starting to be served.

(i) Find the probability that it is more than 5 minutes before she leaves the checkout. [4]

Sally's friend Harry joins a queue at a non-express checkout exactly one minute after Sally joins the express queue. There are 2 other people in the non-express queue, the first of whom is just starting to be served. The time, in seconds, spent at the checkout by a customer in the non-express queue may be assumed to be normally distributed with mean 85 and standard deviation 15.

(ii) Show that the probability that Harry leaves his checkout before Sally leaves her checkout is approximately 0.042. [6]

[Question 7 is printed overleaf.]

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

$$f(x) = \begin{cases} 1 - \frac{1}{2}x & 0 \leqslant x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the (cumulative) distribution function of X. [3]

The random variable Y is defined by $Y = aX + b$, where the constants a and b, with $a > 0$, are chosen so as to make $E(Y) = 0$ and $\text{Var}(Y) = 2$.

(ii) Given that $E(X) = \frac{2}{3}$ and $\text{Var}(X) = \frac{2}{9}$, show that $a = 3$ and find the value of b. [3]

(iii) Find an expression for the (cumulative) distribution function of Y. [4]

(iv) Hence show that the probability density function of Y is given by

$$g(y) = \begin{cases} \frac{2}{9} - \frac{1}{18}y & -2 \leqslant y \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$ [4]

1. Po(4.1)	B1	Summing 2 Poissons, to give Po(4.1)
Probability = 1 – 0.7693	M1	Use of Poisson tables to find $1 - P(X+Y \leq 5)$
= 0.231	A1 3	Any answer rounding to 0.231
---	---	---
2. $P(1) = P(3) = P(4) = P(6) = \frac{1}{6}$, $P(5) = \frac{1}{3}$	B1	Correct probabilities or equivalent.
Expected frequencies are 7.5, 7.5, 7.5, 15, 7..5	B1	All expected frequencies correct
$\sum \frac{(O-E)^2}{E} = \frac{(10-7.5)^2}{7.5} + \frac{(7-7.5)^2}{7.5} + \ldots$	M1	At least one correct term – their E values.
$= 3.13$	A1	Any answer rounding to 3.1
Less than the appropriate chi square value of 7.779	M1	Compare chi square statistic and table value, correct degrees of freedom.
There is insufficient evidence to reject null hypothesis that balance is unaffected – conclude balance not affected by addition of 3 spots.	A1 6	Correct conclusion, in context, after completely correct methods.
SR: $H_0: P(X = 5) = \frac{1}{3}$ ($H_1: P(X = 5) \neq \frac{1}{3}$		
$X \sim B(45, \frac{1}{3})$	B1	
Approximate by N(15,10)		
$z = (18 - 15)/\sqrt{10} = 0.948$	B1	
$0.948 < 1.645$ =>	M1	Compare z statistic with 1.645 or 1.282
Accept H_0: balance is unaffected Max 4/6	A1	

3 (i) The two populations are normally distributed.	B1	
The two populations have equal variance.	B1 2	
(ii) $H_0: \mu_x = \mu_y$ $H_1: \mu_x \neq \mu_y$	B1	Both hypotheses correct
$S_p^2 = (107.58 + 87.52)/17 = 11.476.$	B1	S_p^2 correct to 1 dp.
$t = \pm (31.0 - 35.2) / \sqrt{(11.476(\frac{1}{11} + \frac{1}{8}))}$	M1	Formula for 2 sample t-test – their S_p^2
$= \pm 2.668$	A1	Completely correct expression
	A1	Correct to 1 dp
$\lvert t \rvert = 2.668$ is greater than critical t value 2.110	M1	For correct df = 17 and comparison of t statistic with table t value.
Sufficient evidence to reject H_0 and conclude that the lengths of line drawn by the two groups are different.	A1 7	Correctly stated conclusion, in context, following correct methods including reasonable attempt at 2 sample formula with correct form for variance.
SR (31-35.2)/$\sqrt{(10.578/11+12.502/8)}$ (= 2.634)		
gets B1B0B1B0 (for S_p^2) M1A1(exactly as given) A0		
M1A1(comparison with z=1.96 or t_{17}=2.110) Max 6/9		
---	---	---
4. H_0 : Council Tax band and whether or not they have a double garage are independent of each other.	B1	Null hypothesis stated.
Expected frequencies 20.4 21.9 13.2 4.5	M1	Correct method for at least one cell
47.6 51.1 30.8 10.5	A1	All correct to 1 dp.
Combining last 2 columns gives O E	M1	Combining 3rd and 4th column
28 17.7		
31 41.3		
$\sum \frac{(O-E)^2}{E} = \frac{(13-20.4)^2}{20.4} + \frac{(19-21.9)^2}{21.9} + ..$	M1	At least one correct term – their E values.
$= 2.6843 + 0.3840 + \ldots = 12.946$	A1	Correct to 1 dp.
This is greater than the appropriate chi square value 4.605	M1	c.f. chi-square statistic and table value
Reject H_0 , tax bands and number of garages are not independent.	A1 8	Correct conclusion from comparison, in context, here or in statement of H_0.
SR. No combining B1 M1 A1 M0 M1 A0 M1A1 Max 6/8		Comparison with $\chi^2(3) = 6.251$
Combining 2 cells B1 M1 A1 M0 M1A0 M1A1 Max 6/8		Comparison with $\chi^2(6) = 10.64$, $\chi^2(3) = 6.251$ or $\chi^2(2) = 4.605$

5. (i) $\bar{x}$ = 50.2	B1	At any stage – may be implied
Interval is $50.2 \pm 2.326 \frac{\sqrt{0.19}}{\sqrt{12}}$	M1	Calculation of form $\bar{x} \pm z.\sigma/\sqrt{n}$
	B1	Relevant use of 2.326
Hence $49.91 < \mu < 50.49$	A1 4	Correct to 2 dp – allow [49.91,50.49]
SR $49.8 \pm 2.326 . \sqrt{0.19}/\sqrt{12}$ B0 M1 A1A0		
(ii) s^2 = 0.236(36) (or 0.2166 if $\sqrt{11}$ used subsequently)	B1	Correct unbiased estimate of σ^2
Interval is $50.2 \pm 2.718 \frac{\sqrt{0.236..}}{\sqrt{12}}$	M1	Calcn of form $\bar{x} \pm t\sqrt{}$ (s^2/n), their s^2
	B1	Relevant use of 2.718
Hence $49.82 < \mu < 50.58$	A1 4	Unbiased s^2 reqd. Correct to 2 dp (penalise only once)
SR $49.8 \pm 2.718 . \sqrt{0.236}/\sqrt{12}$ M1 A1A0		
(iii) 50.0 is within this 98% confidence interval	B1	Correct statement from their interval.
So hypothesis test will lead to acceptance of null hypothesis	B1 2	- for each mark.
---	---	---
6. (i) $\Sigma S \sim N(264,200)$	B1B1	Mean correct - `Normal` not reqd. Variance correct - `Normal` not reqd.
$P(\Sigma S > 300) = P(z > \frac{300-264}{\sqrt{200}})$	M1	$\frac{300-\mu}{\sigma}$
$= 1 - \Phi(2.5455)$		
$= 0.0054$ or 0.0055	A1 4	Correct to 2 sf.
(ii) $\Sigma H \sim N(255,675)$	B1	Mean and variance, Normal not required
$\Sigma H - \Sigma S \sim N(255 - 264, \quad)$	M1	Using their means
$\sim N(-9, 875)$	A1	Both correct
$P(\Sigma H - \Sigma S < -60)$ or equivalent	M1	Correct statement, allow – 1 minute.
$= P(z < \frac{-60+9}{\sqrt{875}})$	M1	Their values – not mixed units.
$= 1 - \Phi(1.724) = 0.0423$ [AG]	A1 6	Some working must be given
SR: (i) n =7, B1 B1 M1 A0 Max 3/4		
$\Sigma S \sim N(231,175)$ $P(z > (300-231)/\sqrt{275})$		
(ii) n = 7, B1M1A0 M1 M1A0 Max 4/6		
$\Sigma H - \Sigma S \sim N(255 - 231, \quad)$, $P(z < (-60 + 24)/\sqrt{850})$		
if also m = 2, $\Sigma H \sim N(170,450)$ – B0, then as above.		

7. (i) $F(x) = \int_0^x (1 - \frac{1}{2}x)dx = x - \frac{1}{4}x^2 \quad (0 \leq x \leq 2)$	M1 A1 B1 3	Use of $\int f(x)\,dx$, limits not required
$x < 0 \Rightarrow F(x) = 0$; $x > 2 \Rightarrow F(x) = 1$		
(ii) $\frac{2}{3}a + b = 0$	B1	Use of $E(aX + b) = aE(X) + b$ and
$a^2 \frac{2}{9} = 2.$	M1	$V(aX + b) = a^2 V(X)$
$a = 3.$ and $b = -2$ [AG]	A1 3	Both correct.
(iii) $G(y) = P(Y \leq y) = P(3X - 2 \leq y)$	M1	
$= P(X \leq \frac{1}{3}y + \frac{2}{3})$	A1	For converting $y = 3x-2$ into $x = \frac{1}{3}(y+2)$
$= F(\frac{1}{3}y + \frac{2}{3})$	M1	Use of F(x)
$= \frac{1}{3}y + \frac{2}{3} - \frac{1}{4}(\frac{1}{3}y + \frac{2}{3})^2$	A1 4	
or $= \frac{1}{36}(-y^2 + 8y + 20)$ or equivalent		
(iv) $g(y) = G`(y) = \frac{1}{3} - \frac{1}{2} \cdot \frac{1}{3}(\frac{1}{3}y + \frac{2}{3})$	M1	Differentiation of their G(y) – must be seen
$= \frac{2}{9} - \frac{1}{18}y$ [AG]	A1	
<u>Either</u> Range is $0 \leq \frac{1}{3}y + \frac{2}{3} \leq 2$	M1	Substitution in $0 \leq x \leq 2$
$-2 \leq y \leq 4$ [AG]	A1	
<u>Or</u> $x=2 \Rightarrow y = 3.2 - 2 = 4$	M1	
and $x = 0 \Rightarrow y = 3.0 - 2 = -2$	A1 4	