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 **25 JANUARY 2005** 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.

[Turn over

1 A questionnaire sent to doctors contained an item on smoking. The researcher wished to test whether smoking profile depends on age and she set up a contingency table with the following headings. Data values are omitted.

Smoking Profile

		Never smoked	Current smokers	Ex-smokers
	< 40			
Age	$40 - 50$			
	> 50			

In the test it was found necessary to combine the first two rows. The value of χ^2 was then calculated to be 10.474. Determine the conclusion of the test at the 5% significance level. [4]

2 A hardware shop sells wood screws produced by two manufacturers, A and B. On average, 2% of those produced by A have faulty heads and $2\frac{1}{2}$% of those produced by B have faulty heads. I buy 125 screws produced by A and 100 screws produced by B. The total number of screws with faulty heads is denoted by F. It may be assumed that the screws purchased are random samples.

(i) Find the exact value of $\mathrm{E}(F)$. [2]

(ii) Using suitable Poisson approximations, find the probability that exactly 3 of the 225 screws have faulty heads. [3]

3 The lifetime in years of a particular machine is a continuous random variable T with probability density function given by

$$\mathrm{f}(t) = \begin{cases} \frac{1}{180}t^2 & 0 \leqslant t \leqslant 6, \\ \frac{1}{30}(12 - t) & 6 < t \leqslant 12, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that the expected lifetime is 6.6 years. [4]

(ii) The total running cost for a machine whose lifetime is T years is £$(120 + 0.5T)$. Find the expected value of the total running cost. [2]

4 A new coffee machine was installed in a cafeteria and information was sought regarding the amount of caffeine dispensed in a cup of (low-caffeine) coffee. The amounts of caffeine in a random sample of 80 cups of coffee were measured. These amounts, x mg, are summarised by $\Sigma x = 552$ and $\Sigma x^2 = 3924$.

(i) Find an unbiased estimate of the variance of the amount of caffeine dispensed in a cup. [2]

The mean amount of caffeine dispensed in a cup is μ mg.

(ii) Find a 99% confidence interval for μ. [3]

(iii) State why it is necessary to use the Central Limit Theorem in calculating the interval. [1]

(iv) If the confidence level of 99% is reduced, state whether the resulting confidence interval will be

(a) narrower or wider, [1]

(b) more likely or less likely to contain μ. [1]

5 John cycles to work each day, a distance of 20 km. Owing to varying traffic conditions, the time for the journey varies. The journey time, T hours, is a continuous random variable with (cumulative) distribution function given by

$$\mathrm{F}(t) = \begin{cases} 0 & t < 1, \\ c(3t^2 - t^3 - 2) & 1 \leqslant t \leqslant 2, \\ 1 & t > 2, \end{cases}$$

where c is a positive constant.

(i) Show that $c = \frac{1}{2}$. [1]

(ii) Show that the median journey time is less than $1\frac{1}{2}$ hours. [2]

The probability density function and (cumulative) distribution function of the average speed V km h^{-1} for the journey are denoted by $\mathrm{g}(v)$ and $\mathrm{G}(v)$ respectively.

(iii) Show that $\mathrm{G}(v) = 1 - \mathrm{F}\left(\frac{20}{v}\right)$. [3]

(iv) Hence show that, over the interval $10 \leqslant v \leqslant 20$,

$$\mathrm{G}(v) = 2 - \frac{600}{v^2} + \frac{4000}{v^3}. \quad [2]$$

(v) Find $g(v)$ over the interval $10 \leqslant v \leqslant 20$. [2]

[Questions 6 and 7 are printed overleaf.]

6 In order to test a coin for bias, the following procedure was carried out 128 times.

The coin is tossed repeatedly until a head is obtained.

The score, x, is the number of tosses up to and including the one with the head.

A frequency table of the results is as follows.

x	1	2	3	4	5	6	7	$\geqslant 8$
Frequency	53	28	19	12	8	6	2	0

(i) By fitting a $\text{Geo}(\frac{1}{2})$ distribution, show that there is evidence at the $2\frac{1}{2}\%$ significance level that the coin is biased. [6]

(ii) State, giving your reasons, whether the bias is towards a head or towards a tail. [2]

(iii) Show that the data implies a total of 304 tosses, 128 of which are heads. [2]

(iv) Find an approximate 95% confidence interval for the probability that the coin comes down heads. [3]

7 Type A resistors sold at an electronics store have a nominal resistance of 3.50 ohms and Type B have a nominal resistance of 3.00 ohms. Random samples of 6 Type A and 5 Type B resistors were measured with the following results, in ohms.

A	3.41	3.52	3.38	3.50	3.43	3.46
B	2.90	2.92	2.88	2.97	3.01	

The population mean resistances of Type A and Type B resistors are denoted by μ_A ohms and μ_B ohms respectively. The resistances of both types are normally distributed.

(i) Test at the 5% significance level whether $\mu_A < 3.50$. [6]

(ii) Stating a necessary assumption, find a 95% confidence interval for $\mu_A - \mu_B$. [6]

(iii) State, giving a reason, whether the data is consistent with $\mu_A - \mu_B = 0.50$. [2]

S3 Jan 2005

1. $\nu = (\text{rows}-1)(\text{columns}-1)$
$= (2-1)(3-1)$
$= 2$

$\chi^2_2\ 5\% = 5.991$

H_0: smoking and age are independent
H_1: smoking & age are not independent.

$\chi^2_{calc} = 10.475 > 5.991$

$\therefore$ reject H_0
there is evidence that smoking & age are not independent.

2. A : $X \sim B(125, .02)$ X = no of faulty heads on screws
B : $Y \sim B(100, .025)$ Y = " " " " " "

(i) $E(F) = E(X) + E(Y)$
$= 125 \times .02 + 100 \times .025$
$= 5$

(ii) approxs $X \sim Po(125 \times 0.02)$
$X \sim Po(2.5)$
$\left\{ \begin{array}{l} n \text{ is large} \\ p \text{ is } < 0.1 \\ \& np < 5 \end{array} \right.$

$Y \sim Po(100 \times 0.025)$
$Y \sim P(2.5)$
$\left\{ \text{as above} \right\}$

$F \sim Po(2.5 + 2.5)$
$\sim Po(5)$
$P(F=3) = \dfrac{e^{-5} 5^3}{3!}$
$= 0.1404$ (4dp)

3. $E(T) = \int_0^6 t\left(\dfrac{t^2}{180}\right)dt + \int_6^{12} \dfrac{t(12-t)}{30}\,dt$
$= \dfrac{1}{180}\left[\dfrac{t^4}{4}\right]_0^6 + \dfrac{1}{30}\left[6t^2 - \dfrac{t^3}{3}\right]_6^{12}$
$= 1.8 - 0 + \dfrac{1}{30}\left[288 - 144\right]$
$= 1.8 + 4.8$
$= \underline{6.6 \text{ years}}$ as req'd.

(ii) $C = 120 + 0.5T$

$E(C) = 120 + 0.5\,E(T)$

$= 120 + 0.5 \times 6.6$

$= 120 + 3.3$

$= \underline{£123.3}$ expected total running cost.

4. $n = 80, \quad \Sigma x = 552, \quad \Sigma x^2 = 3924$

(i) $\hat{\sigma}^2 = \frac{n}{n-1}\left[\frac{\Sigma x^2}{n} - \bar{x}^2\right]$

$= \frac{80}{79}\left[\frac{3924}{80} - \left(\frac{552}{80}\right)^2\right]$

$= 1.4582\ldots$

$= \underline{1.458}$ (4sf)

(ii) 99% CI for μ is $\bar{x} \pm 2.576 \frac{\hat{\sigma}}{\sqrt{n}}$ $\quad X$ = mg of caffeine in the coffee

$\frac{552}{80} \pm 2.576 \times \sqrt{\frac{1.458}{80}}$

$6.55 \leq \mu \leq 7.25$

(iii) $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ by CLT, this is appropriate as although underlying dist is not known to be normal $n > 30$ and hence the dist of $\bar{X}$ will be approx normal.

(iv) (a) narrower

(b) less likely to contain μ

5. (i) Know that $F(2) = 1$ $\quad \therefore \quad c(3 \times 2^2 - 2^3 - 2) = 1$

$c[12 - 8 - 2] = 1$

$2c = 1$

$c = \frac{1}{2}$

(ii) $F(1.5) = \frac{1}{2}\left(3 \times (1.5)^2 - (1.5)^3 - 2\right)$

$= 0.6875$

> 0.5

$\therefore$ median must be < 1.5

5. $V = \frac{20}{T}$

Now $G(v) = P(V \le v)$

$= P\left(\frac{20}{T} \le v\right)$

$= P\left(v \ge \frac{20}{T}\right)$

$= 1 - P\left(v \le \frac{20}{T}\right)$

$= 1 - P\left(t \le \frac{20}{v}\right) \quad t > 0, v > 0$

$= 1 - F\left(\frac{20}{v}\right)$

(iv) $t = 1 \quad v = 20$

$t = 2 \quad v = 10.$

$$G(v) = 1 - \frac{1}{2}\left(3 \times \left(\frac{20}{v}\right)^2 - \left(\frac{20}{v}\right)^3 - 2\right)$$

$$= 1 - \frac{3}{2} \times \frac{400}{v^2} + \frac{1}{2} \times \frac{8000}{v^3} + 1$$

$$= 2 - \frac{600}{v^2} + \frac{4000}{v^3}$$ as req'd $\quad 10 \le v \le 20$

(v) $G(v) = 2 - 600v^{-2} + 4000v^{-3}$

$$g(v) = \frac{d(G(v))}{dv} = 1200v^{-3} - 12000v^{-4}$$

$$= \frac{1200}{v^3} - \frac{12000}{v^4} \quad 10 \le v \le 20$$

6.

x	$P(X=x)$	expected frequences $(\times 128)$		O		$\frac{(O-E)^2}{E}$
1	0.5	64		53		1.890625
2	0.25	32		28		0.5
3	0.125	16		19		0.5625
4	0.0625	8		12		2
5	0.03125	4	8	8	16	8
6	0.015625	2		6		
7	0.0078125	1		2		
7,8	0.0078125	1		0		

$\chi^2_{calc} = 12.953125$

$H_0: X \sim Geo(\frac{1}{2})$

$H_1: X \nsim Geo(\frac{1}{2})$ $\quad \nu = 5 - 1 = 4$ $\quad \chi^2_{3}\ 2\frac{1}{2}\% = 11.14$

Conclude $12.953.. > 11.14$ $\therefore$ reject H_0, there is evidence at $2\frac{1}{2}\%$ level that $X \nsim Geo(\frac{1}{2})$, ie the coin is biased.

(ii) the coin seems to be biased towards tails as the observed freq for $X = 1$ & 2 are below expected. These observations account for 0.75 of the probabilities.

(iii) total no of tosses $= \Sigma x f$

$= 53 \times 2\times 28 + 3\times 19 + 4\times 12 \cdots$

$= 53 + 56 + 57 + 48 + 40 + 36 + 14$

$= 304$

(iv) $P_s = \frac{128}{304}$ if p = true prop^n of heads

95% CI for p is

$$P_s \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\frac{128}{304} \pm 1.96\sqrt{\frac{\left(\frac{128}{304}\right)\left(\frac{176}{304}\right)}{304}}$$

$$0.42105 \pm 0.05550\ldots$$

$$0.3656 \leq p \leq 0.4766$$

$$0.37 \leq p \leq 0.48 \quad (2sf)$$

Q7 (i) $H_0: \mu_A = 3.50$

$H_1: \mu_A < 3.50$

$\bar{x} = 3.45$

$\hat{\sigma} = 0.053665\ldots$

$n = 6$

under H_0 $X \sim N(3.50, \sigma^2)$, σ^2 unknown, $\therefore$ use t test

critical value $t_5\ 5\% = -2.015$

test statistic: $t = \frac{\bar{x} - \mu_A}{\hat{\sigma}/\sqrt{n}} = \frac{3.45 - 3.50}{0.053665\ldots/\sqrt{6}}$

$= -2.28$ (3sf).

Conclusion $-2.28 < -2.015$

$\therefore$ reject H_0 at 5%.

evidence for the mean resistance of A < 3.50 ohms

(ii) assume a common variance.

$$\hat{\sigma}^2_{pooled} = \frac{(n_A-1)\hat{\sigma}_A^2 + (n_B-1)\hat{\sigma}_B^2}{n_A+n_B-2}$$

$\bar{x}_A = 3.45$ $n_A = 6$

$\bar{x}_B = 2.936$ $n_B = 5$

$$= \frac{5 \times 0.0028799 + 4 \times 0.00282999}{6+5-2}$$

$$= 0.0028577$$

$$95\%\ CI \text{ for } \mu_A - \mu_B = (\bar{x}_A - \bar{x}_B) \pm t_{9}^{5\%} \sqrt{\hat{\sigma}_p^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}$$

$$= 3.45 - 2.936 \pm 2.262\sqrt{0.0028577\left(\frac{1}{6}+\frac{1}{5}\right)}$$

$$= 0.514 \pm 0.0443376$$

$$0.470 \leq \mu_A - \mu_B \leq 0.558$$

(iii) data is consistent with $\mu_A - \mu_B = 0.5$ as 0.5 is in the 95% CI for $\mu_A - \mu_B$