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 **21 JUNE 2001** 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 4 printed pages.

[Turn over

1 A student tests a 4-sided die to see if it is biased. She rolls the die 100 times and obtains the following frequencies.

Score on die	1	2	3	4
Frequency	32	22	30	16

(i) Show that the value of the appropriate test statistic is 6.56. [2]

(ii) Complete this χ^2 test, using a 10% significance level. [2]

2 In 1978 a County Council carried out a survey of the elm trees in the county to find out what proportion of the elm trees had Dutch Elm disease. In a random sample of 675 elm trees, 432 were found to have the disease. Calculate a 90% confidence interval for the proportion of elm trees in the county that had Dutch Elm disease. [5]

3 A regional water company was considering adding fluoride to the tap-water it supplies. The company commissioned a sample survey of the population in its region. One of the tables in the report on the survey is given below.

Views on fluoridation

	In favour	Against	Total
Men	184	224	408
Women	291	279	570
Total	475	503	978

Test, at the 10% significance level, the hypothesis that, in this region, views on fluoridation are independent of gender. [8]

4 A farmer decided to test the effectiveness of a food additive for cows. She chose a random sample of 12 of her cows and recorded their milk yields over a period of one week. She then included the additive in the cows' food for a period of several weeks and recorded their milk yields again in the last of these weeks. The increases in milk yields, x litres, for the 12 cows are summarised by $\Sigma x = 15.4$ and $\Sigma x^2 = 63.88$.

(i) Use a t-test to test, at the 1% significance level, whether there has been an increase in the mean milk yield. [7]

(ii) State any assumption needed to carry out your test. [1]

5 A commercial grower of blackcurrants wanted to compare the weight of fruit produced by two different varieties of blackcurrant bush. He took a random sample of 14 bushes of the first variety and 13 bushes of the second variety and recorded their yields, in pounds weight, during one summer. The results are summarised below.

$$\text{First variety:} \quad \Sigma x = 205.8, \quad \Sigma(x-\bar{x})^2 = 23.66, \quad n_x = 14.$$
$$\text{Second variety:} \quad \Sigma y = 179.4, \quad \Sigma(y-\bar{y})^2 = 15.73, \quad n_y = 13.$$

(i) Find a 95% confidence interval for the difference between the population mean weights of fruit produced by the two different varieties of blackcurrant bush. (You may assume that the yields of each of the two varieties of blackcurrant bush are normally distributed and that the population variances are equal.) [7]

(ii) Without carrying out any further calculations, state whether, in a two-tail test at the 5% significance level, the hypothesis 'the population mean weights of fruit produced by the two different varieties are equal' would be accepted. Give a reason for your answer. [2]

6 A golfer has played a round of golf on his local golf course every Sunday morning, for 50 weeks of the year, for many years. On average he has lost a golf ball once for every 20 rounds that he has played. His partner, who plays with him every week, has lost a ball, on average, once for every 25 rounds that she has played.

(i) Use appropriate approximations to find the probability that between them they will lose a total of more than 5 balls but fewer than 10 balls in the next year. (You may assume that neither player ever loses more than one ball in a round.) [5]

At the beginning of the year they agree to pay money into a fund to help pay for their fortnight's holiday. The first golfer pays in £100 every time he loses a ball but his partner, because she loses fewer balls, pays in £150 every time she loses a ball.

(ii) Find the expected value of the fund after 50 weeks. [2]

(iii) Find the variance of the value of the fund after 50 weeks. [3]

(iv) State an assumption which you need to make in order for the above calculations to be valid, and state one place where this assumption is used. [2]

[Question 7 is printed overleaf.]

[Turn over

7 The amount of petrol sold by a filling station in a randomly chosen day can be modelled by the random variable X, where X is measured in thousands of litres. The probability density function of X is given by

$$f(x) = \begin{cases} kx & 0 \leqslant x \leqslant 2, \\ k(4-x) & 2 < x \leqslant 3, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Find, in terms of k, the (cumulative) distribution function of X. [5]

(ii) Show that $k = \frac{2}{7}$. [2]

(iii) Find the probability that, in a randomly chosen day, the amount of petrol sold is less than 2500 litres. [2]

(iv) Find the probability that, in a random sample of 7 days, the amount of petrol sold is more than 2500 litres on exactly 5 out of these 7 days. [3]

(v) Say whether the calculation in part **(iv)** would be valid if 7 consecutive days were chosen. Give a reason for your answer. [2]

S3 jun 01

1. H_0: the die is fair, i.e. $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$
H_1: not all the probabilities are equal

o_i	32	22	30	16
e_i	25	25	25	25

$$X^2 = \sum \frac{(o_i - e_i)^2}{e_i} = \frac{7^2}{25} + \frac{3^2}{25} + \frac{5^2}{25} + \frac{9^2}{25}$$

$$= \frac{164}{25} = 6.56$$

$\nu = 4 - 1 = 3$
and the χ_3^2 value for the 10% level is 6.251, so we reject H_0 and conclude that there is significant evidence to suggest that this die is biased.

---

2. 90% CI for proportion is $p_s \pm 1.645\sqrt{\frac{p_s q_s}{n}}$

so $\frac{432}{675} \pm 1.645\sqrt{\frac{\frac{432}{675}\times\frac{243}{675}}{675}}$

and C I is $(0.610, 0.670)$

---

3. H_o: attitude to fluoridation is independent of gender
H_1: it isn't

The e_i's:

198.16	209.84
276.84	293.16

and using Yates' correction,
$$X^2 = \sum \left(\frac{(|o_i - e_i| - 0.5)^2}{e_i}\right)$$
$= 0.94164 + 0.88923 + 0.67402 + 0.63650$
$= 3.14$
with $\nu = 1$, the 10% critical value of χ^2 is 2.706, so we reject H_0
and conclude that there is significant evidence to suggest that views on fluoridation are not independent of gender. The data suggests consistently that women tend to be more in favour.

---

4. H_0: the additive has no effect on milk yield
H_1: the additive increases milk yield

$s^2 = \frac{12}{11}\left(\frac{63.88}{12} - \left(\frac{15.4}{12}\right)^2\right) = 4.0106 \ldots$

so $t_{11} = \frac{1.28333-0}{\frac{\sqrt{4.01}}{12}}$

$= 7.69$

critical value at 1% level is 2.718, so there is clearly very strong evidence of an increase in the yield.

The assumption necessary for the test was that the differences in yield were Normally distributed.

---

5. Pooled estimate of common variance

$= \frac{23.66 + 15.73}{25}$

$= 1.5756$

The t_{25} value for 95% is 2.060, so a symmetric 95% confidence interval for difference of means (first – second) is

$$\frac{205.8}{14} - \frac{179.4}{13} \pm 2.06 \times \sqrt{1.5756\left(\frac{1}{14} + \frac{1}{13}\right)}$$

$= (0.417, 1.38)$ pounds.

No, because the value 0 lies outside the 95% c.i. for difference of means and would therefore be significant in a hypothesis test at the 5% level.

---

6. Number of balls lost by him per year $\sim P(2.5)$ and by her $\sim P(2)$.

total number lost by both in a year $\sim P(4)$

$$P(5 < total < 10) = \frac{4.5^6 e^{-4.5}}{6!} + \frac{4.5^7 e^{-4.5}}{7!} + \frac{4.5^8 e^{-4.5}}{8!} + \frac{4.5^9 e^{-4.5}}{9!}$$

$= 0.280$

$E(100X + 150Y)$

$= 100 \times 2.5 + 150 \times 2$

$= £550$

$V(100X + 150Y)$

$= 100^2 \times 2.5 + 150^2 \times 2$

$= £^2 70000$

Her losing balls needs to be independent of his losing balls. This was used in the variance calculation, otherwise there would be covariance to add in.

7.

(i)

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2}kx^2 & 0 \le x \le 2 \\ k\left(4x - \frac{x^2}{2} - 4\right) & 2 \le x \le 3 \\ 1 & x \ge 3 \end{cases}$$

(ii)

Since $F(3) = 1$, $\quad k\left(12 - \frac{9}{2} - 4\right) = 1$

so $k = \frac{2}{7}$

(iii)

$$F(2.5) = \frac{23}{28}$$

(iv)

$$\binom{7}{5}\left(\frac{5}{28}\right)^5\left(\frac{23}{28}\right)^2 = 0.00257$$

(v)

No, this would include a weekend, for example, and probabilities may well vary day to day.