OCR

RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 2642

Probability & Statistics 2

Tuesday **18 JUNE 2002** 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 3 printed pages and 1 blank page.

1 Sixty people each make two throws with a fair six-sided die. Using a suitable approximation, calculate the probability that at least four of the sixty obtain two sixes. [5]

2 The standard deviation of a random variable F is 12.0. The mean of n independent observations of F is denoted by $\overline{F}$.

(i) Given that the standard deviation of $\overline{F}$ is 1.50, find the value of n. [3]

(ii) For this value of n, state, with justification, what can be said about the distribution of $\overline{F}$. [2]

3 A certain neighbourhood contains many small houses (with small gardens) and a few large houses (with large gardens). A sample survey of all houses is to be carried out in this neighbourhood. A student suggests that the sample could be selected by sticking a pin into a map of the neighbourhood the requisite number of times, while blindfolded.

(i) Give two reasons why this method does not produce a random sample. [2]

(ii) Describe a better method. [3]

4 The random variable G has mean 20.0 and standard deviation σ. It is given that $P(G > 15.0) = 0.6$. Assume that G is normally distributed.

(i) **(a)** Find the value of σ. [4]

(b) Given that $P(G > g) = 0.4$, find the value of g. [2]

(ii) It is known that no values of G are ever negative. State what this tells you about the assumption that G is normally distributed. [1]

5 The proportion of left-handed adults in a country is known to be 15%. It is suggested that for mathematicians the proportion is greater than 15%. A random sample of 12 members of a university mathematics department is taken, and it is found to include five who are left-handed.

(i) Stating your hypotheses, test whether the suggestion is justified, using a significance level as close to 5% as possible. [7]

(ii) In fact the significance test cannot be carried out at a significance level of exactly 5%. State the probability of making a Type I error in the test. [2]

6 On average a motorway police force records one car that has run out of petrol every two days.

(a) **(i)** Using a Poisson distribution, calculate the probability that, in one randomly chosen day, the police force records exactly two cars that have run out of petrol. [3]

(ii) Using a Poisson distribution and a suitable approximation to the binomial distribution, calculate the probability that, in one year of 365 days, there are fewer than 205 days on which the police force records no cars that have run out of petrol. [5]

(b) State an assumption needed for the Poisson distribution to be appropriate in part **(a)**, and explain why this assumption is unlikely to be valid. [2]

7 The mean solubility rating of widgets inserted into beer cans is thought to be 84.0, in appropriate units. A random sample of 50 widgets is taken. The solubility ratings, x, are summarised by

$$n = 50, \qquad \Sigma x = 4070, \qquad \Sigma x^2 = 336\,100.$$

Test, at the 5% significance level, whether the mean solubility rating is less than 84.0. [10]

8 The time, in minutes, for which a customer is prepared to wait on a telephone complaints line is modelled by the random variable X. The probability density function of X is given by

$$f(x) = \begin{cases} \frac{4}{81}x(9 - x^2) & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find $E(X)$. [3]

(ii) **(a)** Show that the value y which satisfies $P(X < y) = \frac{3}{5}$ satisfies

$$5y^4 - 90y^2 + 243 = 0.$$ [3]

(b) Using the substitution $w = y^2$, or otherwise, solve the equation in part **(a)** to find the value of y. [3]

1 No. of people throwing two sixes (X) ~ $B\left(60, \frac{1}{36}\right) \approx Po\left(\frac{5}{3}\right)$

$$p(X \geq 4) \approx 1 - p(X \leq 3) = 1 - 0.9117 = \mathbf{0.0883} \qquad \text{(3 s.f.)}$$

[5]

---

2 (i) $\dfrac{12\cdot 0}{\sqrt{n}} = 1\cdot 50$ $\qquad$ $\boldsymbol{n = 64}$

[3]

(ii) Since n is large, the Central Limit Theorem tells us that the distribution of $\bar{F}$ will be approximately normal.

[2

]

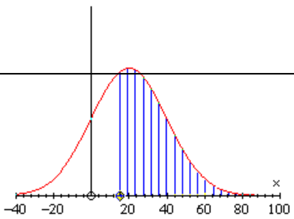
---

3 (i) The houses on large plots are more likely to be chosen than those on small plots. Also, the chosen method is unlikely to give a uniform distribution of pinpricks over the whole neighbourhood. [2]

(ii) A better method would be to assign a number to each house and then use a random number generator to select the houses for the sample. [3]

---

4 $G \sim N\left(20, \sigma^2\right)$



$p(G > 15\cdot 0) = 0\cdot 6$

$\Phi\left(\dfrac{15-20}{\sigma}\right) = 0\cdot 4$

$\Phi\left(\dfrac{5}{\sigma}\right) = 0\cdot 6$ $\qquad$ $\dfrac{5}{\sigma} = 0\cdot 253$ $\qquad$ $\boldsymbol{\sigma = 19\cdot 8}$

[4]

$p(G > g) = 0\cdot 4$ $\qquad$ by symmetry … $\qquad$ $\boldsymbol{g = 25}$

[2]

If G were normally distributed (with the above parameters) then there would be a probability of approximately 16% of taking a negative value. Knowing that G cannot take negative values makes the assumption of normality untenable.

[1]

---

5 $H_0: \ p = 0\cdot 15$ $\qquad H_1: \ p > 0\cdot 15$

On H_0 No. of left-handers $(X) \sim B\left(12,\ 0\cdot 15\right)$ and so $\begin{cases} p(X > 3) = 9\cdot 2\% \\ p(X > 4) = 2\cdot 4\% \end{cases}$

So, using a significance of 2.4%, we reject H_0 if $X>4$.

In the sample taken X=5, and so this provides sufficient evidence on which to reject H_0 and conclude that there are more left-handers amongst mathematicians. [7]

$\text{p}\left(\text{Type I error}\right) = \mathbf{2\cdot 39\%}$ [2]

---

6 (a) On one day, no. of occurrences is **Po(0.5)**

$$\text{p}\left(2 \text{ cars}\right) = e^{-0\cdot 5}\frac{0\cdot 5^2}{2!} = \mathbf{0\cdot 0758} \qquad (3\ s.f.)$$ [3]

No. of days with no cars $\sim \text{B}\left(365,\ e^{-0\cdot 5}\right) \approx N\left(221\cdot 4,\ 87\cdot 11\right)$

$$p(< 205) = \Phi\left(\frac{204\cdot 5 - 221\cdot 4}{\sqrt{87\cdot 11}}\right) = \Phi(-1\cdot 811) = \mathbf{0\cdot 0351} \qquad (3\text{ s.f.})$$ [5]

(b) It's unlikely that the condition of **uniform occurrence** of events is satisfied, because on days of heavy traffic (e.g. holiday time) there are many more vehicles on the road and hence more likelihood of encountering cars that have run out of petrol. [2]

---

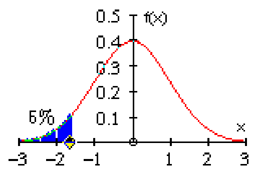
7 H_0: $\mu = 84.0$ $\qquad H_1$: $\mu < 84.0$

Working on H_0, for the sample mean $\overline{X}$ from samples of size 50

$$Z = \frac{\overline{X} - 84}{\sqrt{S^2/50}} \sim \text{N}\left(0,\ 1\right) \qquad \text{where} \quad S^2 = \frac{50}{49}\left(\frac{\sum x^2}{50} - \left(\frac{\sum x}{50}\right)^2\right)$$

So with a one-tail test and a 5% significance level,

we reject H_0 if $Z < -1.645$



For the sample given…..

$$\overline{x} = \frac{4070}{50} = 81\cdot 4 \qquad s^2 = \frac{50}{49}\left(\frac{336100}{50} - (81\cdot 4)^2\right) = 98$$

and $$z = \frac{81.4 - 84}{\sqrt{98/50}} = \mathbf{-1\cdot 857} < -1\cdot 645$$

So significant evidence at the 5% level to conclude that the mean solubility rating is less than 84.0. [10]

8 (i) $\mathrm{E}[X] = \int_0^3 \tfrac{4}{81} x^2 \left(9 - x^2\right) \mathrm{d}x = \tfrac{4}{81}\left[3x^3 - \tfrac{1}{5} x^5\right]_0^3 = \boldsymbol{\tfrac{8}{5}}$ [3]

(ii)

$$\mathrm{p}(X < y) = \tfrac{3}{5} \quad \Rightarrow \quad \int_0^y \tfrac{4}{81} x\left(9 - x^2\right) \mathrm{d}x = \tfrac{3}{5}$$

$$\Rightarrow \quad \tfrac{18}{81} y^2 - \tfrac{1}{81} y^4 = \tfrac{3}{5}$$

$$\Rightarrow \quad \boldsymbol{5y^4 - 90y^2 + 243 = 0} \qquad (\textit{show})$$ [3]

$$\Rightarrow \quad y^2 = \frac{90 \pm \sqrt{3240}}{10} = 3 \cdot 3079\ldots \ or\ 14 \cdot 692\ldots$$

$$\Rightarrow \quad y = \pm 1 \cdot 8187\ldots \text{ or } \pm 3 \cdot 833\ldots$$

But $0 < y < 3$, and so $\boldsymbol{y = 1.82}$ (3 s.f.) [3]