OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 2642

Probability & Statistics 2

Wednesday **23 JUNE 2004** 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 The wingspan of shelduck is normally distributed with mean 121.5 cm and standard deviation 5.3 cm. Find the probability that one randomly chosen shelduck has a wingspan of at least 130 cm. [3]

2 A random sample of 5 people is to be drawn from 10 people whose surnames are Ali, Budd, Cook, Dost, Evans, Fox, Grant, Hall, Ip, Jenks. Use the following extract from a table of random numbers, starting at the beginning, to obtain the sample, and write down the names of the people chosen. Make your method clear.

079 831 130 709 938 423 756 281 787 118 [4]

3 On average 25% of pupils in a large school come to school by bus. Use a suitable approximation to find the probability that, in a random sample of 48 pupils, at least 15 come to school by bus. [6]

4 In a blood test, some blood is placed on a microscope slide and the number of corpuscles in each grid square of the slide is counted. The number of corpuscles per grid square in a sample of blood is a random variable with the distribution $\text{Po}(\mu)$.

(i) For healthy blood, it is known that $\mu = 2.0$. Find the probability that, in a randomly chosen sample of healthy blood, the number of corpuscles counted in one grid square is less than 3. [2]

(ii) A significance test of the null hypothesis $H_0 : \mu = 2.0$ as opposed to the alternative hypothesis $H_1 : \mu < 2.0$ is carried out, using a significance level as close as possible to 5%. The test is based on the total number of corpuscles counted in a group of 4 grid squares.

(a) Find the largest total number of corpuscles counted that would result in rejection of the null hypothesis. You should show the value of any relevant probability. [2]

(b) Given that, in fact, $\mu = 1.75$, find the probability that the test results in a Type II error. [3]

5 The random variable X has the distribution $N(\mu, \sigma^2)$. It is given that $P(X > 2\mu) = 0.0228$.

(i) Find the value of μ in terms of σ. [3]

In order to calculate the actual values of μ and σ, more information is required.

(ii) Explain why neither of the following extra pieces of information would enable you to work out the actual values of μ and σ:

(a) $P(X < 0) = 0.0228$; [1]

(b) $P(X < \mu) = 0.5$. [1]

(iii) Given that $P(X < 7.0) = 0.7881$, calculate the actual values of μ and σ. [4]

6 Requests for the services of a professional services company are received at a constant average rate of 38 per week, independently of one another. If more than 50 requests are received in any one week, the company has to take on extra staff.

(i) Use a suitable approximation to show that the probability that the company has to take on extra staff in one randomly chosen week is 0.021, correct to 3 decimal places. [5]

(ii) Use a suitable approximation to find the probability that, in a year of 50 working weeks, the company has to take on extra staff in more than 2 weeks. [5]

7 The random variable T represents the number of minutes that a train is late for a particular scheduled journey on a randomly chosen day.

(i) Give a reason why T could probably not be well modelled by a normal distribution. [1]

(ii) The following probability density function is proposed as a model for the distribution of T:

$$\mathrm{f}(t) = \begin{cases} \frac{1}{67\,500}t(t-30)^2 & 0 \leqslant t \leqslant 30, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of $\mathrm{E}(T)$. [3]

On a randomly chosen day I will allow for the train to be up to t_0 minutes late. I wish to choose the value of t_0 for which the probability that the train is less than t_0 minutes late is 0.95.

(b) Show that t_0 satisfies the equation

$$t_0^4 - 80t_0^3 + 1800t_0^2 = 256\,500. \qquad [4]$$

(c) Show that the value of t_0 lies between 22 and 23. [2]

8 A psychologist is testing the effect of background music on students' work. She knows that the average time for a randomly chosen student to complete a particular task in the absence of background music is 38.5 seconds. A sample of 50 students took the test with background music being played. The times taken, t seconds, are summarised as follows.

$$n = 50, \ \Sigma t = 1967, \ \Sigma t^2 = 77\,959.$$

(i) Test, at the 5% significance level, whether the presence of background music has an effect on the times taken by students to complete the task. State your hypotheses clearly. [10]

(ii) Give a reason why it is necessary to use the Central Limit Theorem in carrying out your test. [1]

OCR Probability & Statistics 2 June 2004

1 $W \sim \mathrm{N}\left(121 \cdot 5,\ 5 \cdot 3^2\right)$ $\quad \mathrm{p}(W \geq 130) = 1 - \Phi\left(\dfrac{130 - 121 \cdot 5}{5 \cdot 3}\right) = 1 - 0 \cdot 9456 = \mathbf{0 \cdot 0544}$ (3 s.f.) [3]

---

2 one possible method …

$0 \to A, \quad 1 \to B,$ etc. then $07983 \ \to \ \boldsymbol{AHJID}$

[2]

---

3 no. arriving by bus $X \sim \mathrm{B}\left(48,\ 0 \cdot 25\right) \approx \mathrm{N}\left(12,\ 9\right)$

$\mathrm{p}(X \geq 15) = 1 - \Phi\left(\dfrac{14 \cdot 5 - 12}{3}\right) = 1 - \Phi\left(0 \cdot 833\dot{3}\right) = 1 - 0 \cdot 7977 = \mathbf{0 \cdot 202}$ (3 s.f.)

[6]

---

4 $X \sim \mathrm{Po}(2)$ $\quad \mathrm{p}(X < 3) = \mathrm{p}(X \leq 2) = \mathbf{0 \cdot 6767}$ (tables) [2]

$\mathrm{H}_0 : \mu = 2$ $\quad \mathrm{H}_1 : \mu < 2$

If H_0 is true, the no. of corpuscles in 4 squares $X \sim \mathrm{Po}(8)$ and $\begin{cases} \mathrm{p}(X \leq 4) = 0 \cdot 0996 \\ \mathrm{p}(X \leq 3) = 0 \cdot 0424 \end{cases}$

…. so for 5% significance level we will **reject** $\mathbf{H_0}$ **if** $\boldsymbol{X \leq 3}$

[2]

If in fact $\mu = 1.75$, then $\mathrm{p}\left(\text{Type II error}\right) = \mathrm{p}\left(X > 3 \,\middle|\, \lambda = 7\right) = 1 - 0 \cdot 0818 = \mathbf{0 \cdot 918}$ (3 s.f.)

[3]

---

5 $X \sim \mathrm{N}\left(\mu,\ \sigma^2\right)$ $\quad \mathrm{p}(X > 2\mu) = 0 \cdot 0228 \quad \Rightarrow \quad \Phi\left(\dfrac{2\mu - \mu}{\sigma}\right) = 0 \cdot 9772 \quad \Rightarrow \quad \dfrac{\mu}{\sigma} = 2 \cdot 000$

$\boldsymbol{\mu = 2\sigma}$

[3]

$\mathrm{p}(X < 0) = 0.0228$ tells us nothing new … **equivalent to original** (LH tail rather than RH tail)
$\mathrm{p}(X < \mu) = 0 \cdot 5$ gives us no information since it's **true for any normal** distribution.

[2]

$\mathrm{p}(X < 7 \cdot 0) = 0 \cdot 7881 \quad \Rightarrow \quad \Phi\left(\dfrac{7 - \mu}{\sigma}\right) = 0 \cdot 7881 \quad \Rightarrow \quad \dfrac{7 - \mu}{\sigma} = 0 \cdot 800 \quad \Rightarrow \quad \mu + 0 \cdot 8\sigma = 7$

solving the simultaneous equations gives $\boldsymbol{\mu = 5,\ \sigma = 2 \cdot 5}$

[4]

6 no. of requests $X \sim \mathrm{Po}(38) \approx \mathrm{N}\left(38,\ 38\right)$

$$\mathrm{p}(\text{extra staff needed}) = \mathrm{p}(X > 50) = 1 - \Phi\left(\frac{50 \cdot 5 - 38}{\sqrt{38}}\right) = 1 - \Phi(2 \cdot 03) = 1 - 0 \cdot 9788 = \mathbf{0 \cdot 021} \text{ (show)}$$ [5]

no. of weeks in which extra staff are needed $Y \sim \mathrm{B}\left(50,\ 0 \cdot 021\right) \approx \mathrm{Po}(1 \cdot 05)$

$$\mathrm{p}(Y > 2) = 1 - \mathrm{e}^{-1.05}\left(1 + \tfrac{1.05}{1!} + \tfrac{1.05^2}{2!}\right) = \mathbf{0 \cdot 0897} \qquad \text{(3 s.f.)}$$

[5]

---

7 The random variable T is likely to have a considerable positive skew, whereas a normal variable is perfectly symmetric. [1]

$$\mathrm{E}[T] = \tfrac{1}{67\,500}\int_0^{30} t^2 (t-30)^2 \,\mathrm{d}t = \tfrac{1}{67\,500}\int_0^{30}\left(t^4 - 60t^3 + 900t^2\right)\mathrm{d}t = \tfrac{1}{67\,500}\left[\tfrac{1}{5}t^5 - 15t^4 + 300t^3\right]_0^{30} = \mathbf{12}$$

[3]

$\mathrm{p}\left(T < t_0\right) = 0.95 \quad \Rightarrow$

$$\tfrac{1}{67\,500}\int_0^{t_0}\left(t^3 - 60t^2 + 900t\right)\mathrm{d}t = 0 \cdot 95$$

$$\tfrac{1}{4}t_0^{\,4} - 20t_0^{\,3} + 450t_0 = 64\,125$$

$$t_0^{\,4} - 80t_0^{\,3} + 1\,800t_0 = 256\,500 \qquad \text{(show)}$$ [4]

$$\left.\begin{aligned} 22^4 - 80(22)^3 + 1\,800(22) &= 253\,616 \\ 23^4 - 80(23)^3 + 1\,800(23) &= 258\,681 \end{aligned}\right\}$$ **so t_0 lies between 22 and 23**

[2]

---

8 $\mathrm{H}_0 : \mu = 38 \cdot 5 \qquad \mathrm{H}_1 : \mu \neq 38 \cdot 5$

On H_0 …. $Z = \dfrac{\overline{T} - 38 \cdot 5}{\sqrt{\hat{\sigma}^2/50}} \sim \mathrm{N}\left(0,\ 1\right)$ and we reject H_0 when $\boldsymbol{|Z| > 1 \cdot 96}$

for this sample …

$$\left.\begin{aligned} \bar{t} &= \tfrac{1967}{50} = 39 \cdot 34 \\ \hat{\sigma}^2 &= \tfrac{50}{49}\left(\tfrac{77\,959}{50} - 39 \cdot 34^2\right) = \tfrac{50}{49}(11 \cdot 54) = 11 \cdot 78 \end{aligned}\right\} \quad \text{giving} \quad z = \frac{39 \cdot 34 - 38 \cdot 5}{\sqrt{11 \cdot 78/50}} = \mathbf{1 \cdot 73}$$

Since 1.73 < 1.96 there is insufficient evidence on which to reject H_0 and **we conclude that the background music does not make a difference.** [10]

There is no indication that the distribution of work time is normally distributed, so the Central Limit Theorem is relied on to guarantee the normality of $\overline{T}$

[1]