Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Advanced Level

Monday 11 June 2007 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Green)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S2), the paper reference (6684), your surname, other name and signature.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
This paper has 8 questions.
The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N26119A

1. A string AB of length 5 cm is cut, in a random place C, into two pieces. The random variable X is the length of AC.

(*a*) Write down the name of the probability distribution of X and sketch the graph of its probability density function.

(3)

(*b*) Find the values of $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

(3)

(*c*) Find $\mathrm{P}(X > 3)$.

(1)

(*d*) Write down the probability that AC is 3 cm long.

(1)

---

2. Bacteria are randomly distributed in a river at a rate of 5 per litre of water. A new factory opens and a scientist claims it is polluting the river with bacteria. He takes a sample of 0.5 litres of water from the river near the factory and finds that it contains 7 bacteria. Stating your hypotheses clearly test, at the 5% level of significance, the claim of the scientist.

(7)

---

3. An engineering company manufactures an electronic component. At the end of the manufacturing process, each component is checked to see if it is faulty. Faulty components are detected at a rate of 1.5 per hour.

(*a*) Suggest a suitable model for the number of faulty components detected per hour.

(1)

(*b*) Describe, in the context of this question, two assumptions you have made in part (*a*) for this model to be suitable.

(2)

(*c*) Find the probability of 2 faulty components being detected in a 1 hour period.

(2)

(*d*) Find the probability of at least one faulty component being detected in a 3 hour period.

(3)

---

4. A bag contains a large number of coins:

75% are 10p coins,

25% are 5p coins.

A random sample of 3 coins is drawn from the bag.

Find the sampling distribution for the median of the values of the 3 selected coins.

(7)

---

5. (*a*) Write down the conditions under which the Poisson distribution may be used as an approximation to the Binomial distribution.

(2)

A call centre routes incoming telephone calls to agents who have specialist knowledge to deal with the call. The probability of the caller being connected to the wrong agent is 0.01.

(*b*) Find the probability that 2 consecutive calls will be connected to the wrong agent.

(2)

(*c*) Find the probability that more than 1 call in 5 consecutive calls are connected to the wrong agent.

(3)

The call centre receives 1000 calls each day.

(*d*) Find the mean and variance of the number of wrongly connected calls.

(3)

(*e*) Use a Poisson approximation to find, to 3 decimal places, the probability that more than 6 calls each day are connected to the wrong agent.

(2)

---

6. Linda regularly takes a taxi to work five times a week. Over a long period of time she finds the taxi is late once a week. The taxi firm changes her driver and Linda thinks the taxi is late more often. In the first week, with the new driver, the taxi is late 3 times. You may assume that the number of times a taxi is late in a week has a Binomial distribution.

Test, at the 5% level of significance, whether or not there is evidence of an increase in the proportion of times the taxi is late. State your hypotheses clearly.

(7)

---

7. (*a*) (i) Write down two conditions for $X \sim \text{Bin}(n, p)$ to be approximated by a normal distribution $Y \sim \text{N}(\mu, \sigma^2)$.

(2)

(ii) Write down the mean and variance of this normal approximation in terms of n and p.

(2)

A factory manufactures 2000 DVDs every day. It is known that 3% of DVDs are faulty.

(*b*) Using a normal approximation, estimate the probability that at least 40 faulty DVDs are produced in one day.

(5)

The quality control system in the factory identifies and destroys every faulty DVD at the end of the manufacturing process. It costs £0.70 to manufacture a DVD and the factory sells non-faulty DVDs for £11.

(*c*) Find the expected profit made by the factory per day.

(3)

---

8. The continuous random variable X has probability density function given by

$$\text{f}(x) = \begin{cases} \frac{1}{6}x & 0 < x \leq 3 \\ 2 - \frac{1}{2}x & 3 < x < 4 \\ 0 & \text{otherwise} \end{cases}$$

(*a*) Sketch the probability density function of X.

(3)

(*b*) Find the mode of X.

(1)

(*c*) Specify fully the cumulative distribution function of X.

(7)

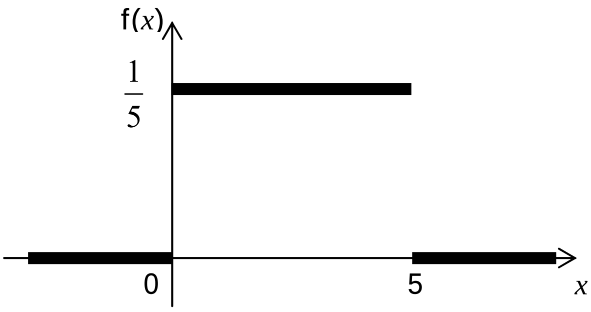
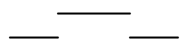
(*d*) Using your answer to part (*c*), find the median of X.

(3)

TOTAL FOR PAPER: 75 MARKS

END

June 2007
6684 Statistics S2
Mark Scheme

Question Number	Scheme	Marks
1(a)	<u>Continuous</u> <u>uniform</u> distribution *or* <u>rectangular</u> distribution.	B1
	 	
	1/5,(0),5 — 0 may be implied by start at y axis	B1
		B1
		(3)
(b)	$E(X) = 2.5$ ft from their a and b, must be a number	B1ft
	$\text{Var}(X) = \frac{1}{12}(5-0)^2$ or attempt to use $\int_0^5 f(x)x^2 dx - \mu^2$ use their f(x)	M1
	$= \frac{25}{12}$ or 2.08 o.e awrt 2.08	A1
		(3)
(c)	$P(X > 3) = \frac{2}{5} = 0.4$ 2 times their 1/5 from diagram	B1ft
		(1)
(d)	$P(X = 3) = 0$	B1
		(1)
		(Total 8)

Question Number	Scheme	Marks
2	One tail test Method 1 $H_o : \lambda = 5$ ($\lambda = 2.5$) may use λ or μ $H_1 : \lambda > 5$ ($\lambda > 2.5$) $X \sim \text{Po}(2.5)$ may be implied $P(X \geq 7) = 1 - P(X \leq 6)$ | [$P(X \geq 5) = 1 - 0.8912 = 0.1088$] att $P(X \geq 7)$ | $P(X \geq 6)$ $= 1 - 0.9858$ | $P(X \geq 6) = 1 - 0.9580 = 0.0420$ $= 0.0142$ | CR $X \geq 6$ awrt 0.0142 $0.0142 < 0.05$ | $7 \geq 6$ or 7 is in critical region or 7 is significant (Reject H_0.) There is significant evidence at the 5% significance level that the factory is polluting the river with bacteria. **or** The scientists claim is justified	B1 B1 M1 M1 A1 M1 B1 (7) Total 7
	Method 2 $H_o : \lambda = 5$ ($\lambda = 2.5$) may use λ or μ $H_1 : \lambda > 5$ ($\lambda > 2.5$) $X \sim \text{Po}(2.5)$ may be implied $P(X < 7)$ | [$P(X < 5) = 0.8912$] att $P(X < 7)$ | $P(X < 6)$ | $P(X < 6) = 0.9580$ $= 0.9858$ | CR $X \geq 6$ wrt 0.986 $0.9858 > 0.95$ | $7 \geq 6$ or 7 is in critical region or 7 is significant (Reject H_0.) There is significant evidence at the 5% significance level that the factory is polluting the river with bacteria. **or** The scientists claim is justified	B1 B1 M1 M1 A1 M1 B1 (7)

	<u>Two tail test</u> <u>Method 1</u> $H_o : \lambda = 5$ ($\lambda = 2.5$) may use λ or μ $H_1 : \lambda \neq 5$ ($\lambda \neq 2.5$) $X \sim \text{Po}(2.5)$ $P(X \geq 7) = 1 - P(X \leq 6)$ | $[P(X \geq 6) = 1 - 0.9580 = 0.0420]$ att $P(X \geq 7)$ | $P(X \geq 7)$ $= 1 - 0.9858$ | $P(X \geq 7) = 1 - 0.9858 = 0.0142$ $= 0.0142$ | CR $X \geq 7$ awrt 0.0142 $0.0142 < 0.025$ | $7 \geq 7$ or 7 is in critical region or 7 is significant (Reject H_0.) There is significant evidence at the 5% significance level that the factory <u>is polluting the river</u> with bacteria. **<u>or</u>** The scientists claim is justified	B1 B0 M1 M1 A1 M1 B1	(7)
	<u>Method 2</u> $H_o : \lambda = 5$ ($\lambda = 2.5$) may use λ or μ $H_1 : \lambda \neq 5$ ($\lambda \neq 2.5$) $X \sim \text{Po}(2.5)$ $P(X < 7)$ | $[P(X < 6) = 0.9580]$ att $P(X < 7)$ | $P(X < 7)$ | $P(X < 7) = 0.9858$ $= 0.9858$ | CR $X \geq 7$ awrt 0.986 $0.9858 > 0.975$ | $7 \geq 7$ or 7 is in critical region or 7 is significant (Reject H_0.) There is significant evidence at the 5% significance level that the factory <u>is polluting the river</u> with bacteria. **<u>or</u>** The scientists claim is justified	B1 B0 M1 M1A1 M1 B1	(7)

Question Number	Scheme		Marks
3(a)	$X \sim \text{Po}\,(1.5)$	need Po and 1.5	B1 (1)
(b)	<u>Faulty</u> components occur at a constant rate. <u>Faulty</u> components occur independently or randomly. <u>Faulty</u> components occur singly.	any two of the 3 only need faulty once	B1 B1 (2)
(c)	$P(X = 2) = \quad P(X \le 2) - P(X \le 1) \qquad \text{or} \qquad \dfrac{e^{-1.5}(1.5)^2}{2}$		M1
	$= 0.8088 - 0.5578$		
	$= 0.251$	awrt 0.251	A1 (2)
(d)	$X \sim \text{Po}(4.5)$	4.5 may be implied	B1
	$P(X \ge 1) = \quad 1 - P(X = 0)$		M1
	$= \quad 1 - e^{-4.5}$		
	$= \quad 1 - 0.0111$		
	$= \quad 0.9889$	awrt 0.989	A1 (3)
			Total 8

Question Number	Scheme		Marks
4	Attempt to write down combinations	at least one seen	M1
	(5,5,5), (5,5,10) any order (10,10,5) any order, (10,10,10)		A1
	(5,10,5), (10,5,5), (10,5,10), (5,10,10),	all 8 cases considered. May be implied by 3 * (10,5,10) and 3 * (5,5,10)	A1
	median 5 and 10		B1
	Median = 5 P(M = m) = $\left(\frac{1}{4}\right)^3 + 3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) = \frac{10}{64}$ = 0.15625	add at least two prob using ¼ and ¾. identified by having same median of 5 or 10 Allow no 3 for M	M1 A1
	Median = 10 P(M = m) = $\left(\frac{3}{4}\right)^3 + 3\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right) = \frac{54}{64}$ = 0.84375		A1
			(7) Total 7

Question Number	Scheme	Marks
5(a)	If $X \sim B(n,p)$ and n is large, $n > 50$ p is small, $p < 0.2$ then X can be approximated by $Po(np)$	 B1 B1 (2)
(b)	P(2 consecutive calls) $= 0.01^2$ $= 0.0001$	M1 A1 (2)
(c)	$X \sim B(5, 0.01)$ may be implied $P(X > 1) = 1 - P(X = 1) - P(X = 0)$ $= 1 - 5(0.01)(0.99)^4 - (0.99)^5$ $= 1 - 0.0480298\ldots - 0.95099\ldots$ $= 0.00098$ awrt 0.00098	B1 M1 A1 (3)
(d)	$X \sim B(1000, 0.01)$ may be implied by correct mean and variance Mean $= np = 10$ Variance $= np(1-p) = 9.9$	B1 B1 B1 (3)
(e)	$X \sim Po(10)$ $P(X > 6) = 1 - P(X \le 6)$ $= 1 - 0.1301$ $= 0.8699$ awrt 0.870	 M1 A1 (2)
		Total 12

Question Number	Scheme	Marks
6	<u>One tail test</u> <u>Method 1</u> $H_o : p = 0.2$ $H_1 : p > 0.2$	 B1 B1
	$X \sim B(5, 0.2)$ may be implied	M1
	$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.9421$ | $[P(X \geq 3) = 1 - 0.9421 = 0.0579]$ $P(X \geq 4) = 1 - 0.9933 = 0.0067$ — att $P(X \geq 3)$ | $P(X \geq 4)$	M1
	$= 0.0579$ | CR $X \geq 4$ — awrt 0.0579	A1
	$0.0579 > 0.05$ | $3 \leq 4$ or 3 is not in critical region or 3 is not significant	M1
	(Do not reject H_0.) There is insufficient evidence at the 5% significance level that there is an increase in the number of times <u>the taxi/driver is late.</u> **Or** Linda's claim is not justified	B1 (7) Total 7
	<u>Method 2</u> $H_o : p = 0.2$ $H_1 : p > 0.2$	 B1 B1
	$X \sim B(5, 0.2)$ may be implied	M1
	$P(X < 3) =$ | $[P(X < 3) = 0.9421]$ $P(X < 4) = 0.9933$ — att $P(X < 3)$ | $P(X < 4)$	
	0.9421 | CR $X \geq 4$ — awrt 0.942	M1A1
	$0.9421 < 0.95$ | $3 \leq 4$ or 3 is not in critical region or 3 is not significant	M1
	(Do not reject H_0.) There is insufficient evidence at the 5% significance level that there is an increase in the number of times the <u>taxi/driver is late.</u> **Or** Linda's claim is not justified	B1 (7)

	Two tail test Method 1 $H_o : p = 0.2$ $H_1 : p \neq 0.2$ $X \sim X \sim B(5, 0.2)$ may be implied $P(X \geq 3) = 1 - P(X \leq 2)$ | $[P(X \geq 3) = 1 - 0.9421 = 0.0579]$ att $P(X \geq 3)$ | $P(X \geq 4)$ $= 1 - 0.9421$ | $P(X \geq 4) = 1 - 0.9933 = 0.0067$ $= 0.0579$ | CR $X \geq 4$ awrt 0.0579 $0.0579 > 0.025$ | $3 \leq 4$ or 3 is not in critical region or 3 is not significant (Do not reject H_0.) There is insufficient evidence at the 5% significance level that there is an increase in the number of times the taxi/driver is late. **Or** Linda's claim is not justified	B1 B0 M1 M1 A1 M1 B1	 (7)
	Method 2 $H_o : p = 0.2$ $H_1 : p \neq 0.2$ $X \sim X \sim B(5, 0.2)$ may be implied $P(X < 3) =$ | $[P(X < 3) = 0.9421]$ att $P(X < 3)$ | $P(X < 4)$ | $P(X < 4) = 0.9933$ 0.9421 | CR $X \geq 4$ awrt 0.942 $0.9421 < 0.975$ | $3 \leq 4$ or 3 is not in critical region or 3 is not significant Do not reject H_0. There is insufficient evidence at the 5% significance level that there is an increase in the number of times the taxi/driver is late. **Or** Linda's claim is not justified Special Case If they use a probability of $\frac{1}{7}$ throughout the question they may gain B1 B1 M0 M1 A0 M1 B1. NB they must attempt to work out the probabilities using $\frac{1}{7}$	B1 B0 M1 M1A1 M1 B1	 (7)

Question Number	Scheme	Marks
7(a) i	If $X \sim B(n,p)$ and n is large or $n > 10$ **or** $np > 5$ **or** $nq > 5$ p is close to 0.5 **or** $nq > 5$ <u>and</u> $np > 5$ then X can be approximated by $N(np, np(1-p))$	B1 B1 (2)
ii	mean = np variance = $np(1-p)$ must be in terms of p	B1 B1 (2)
(b)	$X \sim N(60, 58.2)$ or $X \sim N(60, 7.63^2)$ 60, 58.2 $P(X \geq 40) = P(X > 39.5)$ using 39.5 or 40.5 $= 1 - P\left(z < \pm\left(\frac{39.5-60}{\sqrt{58.2}}\right)\right)$ standardising 39.5 or 40 or 40.5 and their μ and σ $= 1 - P(z < -2.68715\ldots)$ $= 0.9965$ allow answers in range 0.996 – 0.997	B1, B1 M1 M1 A1dep on both M (5)
(c)	$E(X) = 60$ may be implied or ft from part (b) Expected profit = $(2000 - 60) \times 11 - 2000 \times 0.70$ = £19 940.	B1ft M1 A1 (3) Total 12