Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Advanced

Wednesday 21 January 2009 – 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 formulae 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.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions on this paper. 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.

N30996A

1. A botanist is studying the distribution of daisies in a field. The field is divided into a number of equal sized squares. The mean number of daisies per square is assumed to be 3. The daisies are distributed randomly throughout the field.

Find the probability that, in a randomly chosen square there will be

(*a*) more than 2 daisies,

(3)

(*b*) either 5 or 6 daisies.

(2)

The botanist decides to count the number of daisies, x, in each of 80 randomly selected squares within the field. The results are summarised below

$$\sum x = 295 \qquad \sum x^2 = 1386$$

(*c*) Calculate the mean and the variance of the number of daisies per square for the 80 squares. Give your answers to 2 decimal places.

(3)

(*d*) Explain how the answers from part (*c*) support the choice of a Poisson distribution as a model.

(1)

(*e*) Using your mean from part (*c*), estimate the probability that exactly 4 daisies will be found in a randomly selected square.

(2)

---

2. The continuous random variable X is uniformly distributed over the interval $[-2, 7]$.

(*a*) Write down fully the probability density function $\mathrm{f}(x)$ of X.

(2)

(*b*) Sketch the probability density function $\mathrm{f}(x)$ of X.

(2)

Find

(*c*) $\mathrm{E}(X^2)$,

(3)

(*d*) $\mathrm{P}(-0.2 < X < 0.6)$.

(2)

---

3. A single observation x is to be taken from a Binomial distribution $B(20, p)$.

This observation is used to test $H_0 : p = 0.3$ against $H_1 : p \neq 0.3$.

(*a*) Using a 5% level of significance, find the critical region for this test. The probability of rejecting either tail should be as close as possible to 2.5%.
(3)

(*b*) State the actual significance level of this test.
(2)

The actual value of x obtained is 3.

(*c*) State a conclusion that can be drawn based on this value, giving a reason for your answer.
(2)

---

4. The length of a telephone call made to a company is denoted by the continuous random variable T. It is modelled by the probability density function

$$f(x) = \begin{cases} kt, & 0 \leq t \leq 10 \\ 0, & \text{otherwise.} \end{cases}$$

(*a*) Show that the value of k is $\frac{1}{50}$.
(3)

(*b*) Find $P(T > 6)$.
(2)

(*c*) Calculate an exact value for $E(T)$ and for $Var(T)$.
(5)

(*d*) Write down the mode of the distribution of T.
(1)

It is suggested that the probability density function, $f(t)$, is not a good model for T.

(*e*) Sketch the graph of a more suitable probability density function for T.
(1)

---

5. A factory produces components of which 1% are defective. The components are packed in boxes of 10. A box is selected at random.

(*a*) Find the probability that the box contains exactly one defective component.
(2)

(*b*) Find the probability that there are at least 2 defective components in the box.
(3)

(*c*) Using a suitable approximation, find the probability that a batch of 250 components contains between 1 and 4 (inclusive) defective components.
(4)

---

6. A web server is visited on weekdays, at a rate of 7 visits per minute. In a random one minute on a Saturday the web server is visited 10 times.

(*a*) (i) Test, at the 10% level of significance, whether or not there is evidence that the rate of visits is greater on a Saturday than on weekdays. State your hypotheses clearly.

(ii) State the minimum number of visits required to obtain a significant result.

(7)

(*b*) State an assumption that has been made about the visits to the server.

(1)

In a random two minute period on a Saturday the web server is visited 20 times.

(*c*) Using a suitable approximation, test at the 10% level of significance, whether or not the rate of visits is greater on a Saturday.

(6)

---

7. A random variable X has probability density function given by

$$f(x) = \begin{cases} -\frac{2}{9}x + \frac{8}{9}, & 1 \leq x \leq 4 \\ 0, & \text{otherwise.} \end{cases}$$

(*a*) Show that the cumulative distribution function F(x) can be written in the form $ax^2 + bx + c$, for $1 \leq x \leq 4$ where a, b and c are constants.

(3)

(*b*) Define fully the cumulative distribution function F(x).

(2)

(*c*) Show that the upper quartile of X is 2.5 and find the lower quartile.

(6)

Given that the median of X is 1.88,

(*d*) describe the skewness of the distribution. Give a reason for your answer.

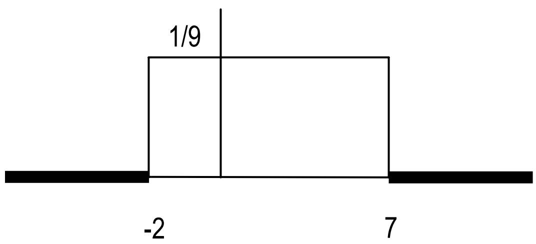
(2)

TOTAL FOR PAPER: 75 MARKS

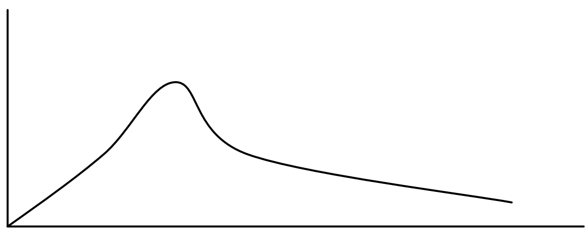
END

January 2009
6684 Statistics S2
Mark Scheme

Question Number	Scheme	Marks
1	The random variable X is the number of daisies in a square.	
	Poisson(3)	B1
(a)	$1 - P(X \le 2) = 1 - 0.4232$ $1 - e^{-3}(1 + 3 + \frac{3^2}{2!})$	M1
	$= 0.5768$	A1
		(3)
(b)	$P(X \le 6) - P(X \le 4) = 0.9665 - 0.8153$ $e^{-3}\left(\frac{3^5}{5!} + \frac{3^6}{6!}\right)$	M1
	$= 0.1512$	A1
		(2)
(c)	$\mu = 3.69$	B1
	$\text{Var}(X) = \frac{1386}{80} - \left(\frac{295}{80}\right)^2$	M1
	$= 3.73/3.72/3.71$ accept $s^2 = 3.77$	A1
		(3)
(d)	For a Poisson model, Mean = Variance ; For these data $3.69 \approx 3.73$	
	$\Rightarrow$ Poisson model	B1
		(1)
(e)	$\frac{e^{-3.6875} 3.6875^4}{4!} = 0.193$ allow their mean or var	M1
	Awrt 0.193 or 0.194	A1 ft
		(2)

Question Number	Scheme	Marks
2 (a)	$f(x) = \begin{cases} \frac{1}{9} & -2 \le x \le 7 \\ 0 & otherwise \end{cases}$	B1 B1 (2)
(b)		B1 B1 (2)
(c)	E(X) = <u>2.5</u> Var (X) = $\frac{1}{12}(7+2)^2$ or <u>6.75</u> both	B1
	E (X^2) = Var (X) + E(X)^2	M1
	= 6.75 + 2.5^2 = 13	A1
	alternative	(3)
	$\int_{-2}^{7} x^2 f(x)dx = \left[\frac{x^3}{27}\right]_{-2}^{7}$ $\int$ x^2"f(x)" attempt to integrate and use limits of -2 and 7	B1 M1
	= 13	A1
(d)	P(-0.2 < X < 0.6) = $\frac{1}{9} \times 0.8$	M1
	= $\frac{4}{45}$ or 0.0889 0r equiv awrt 0.089	A1
		(2)

Question Number	Scheme	Marks
3 (a)	$X \sim B(20, 0.3)$	M1
	$P(X \leq 2) = 0.0355$	
	$P(X \geq 11) = 1 - 0.9829 = 0.0171$	
	Critical region is $(X \leq 2) \cup (X \geq 11)$	A1 A1 (3)
(b)	Significance level = 0.0355 + 0.0171, = 0.0526 or 5.26%	M1 A1 (2)
(c)	Insufficient evidence to reject H_0 **Or** sufficient evidence to accept H_0 /not significant	B1 ft
	$x = 3$ (or the value) is not in the critical region or 0.1071> 0.025	B1 ft (2)
	Do not allow inconsistent comments	

Question Number	Scheme	Marks
4 (a)	$\int_0^{10} kt\mathrm{d}t = 1$ or Area of triangle = 1	M1
	$\left[\frac{kt^2}{2}\right]_0^{10} = 1$ or 10 x0.5 x 10k =1 or linear equation in k	M1
	$50k = 1$	
	$k = \frac{1}{50}$ cso	A1
		(3)
(b)	$\int_6^{10} ktdt = \left[\frac{\mathrm{kt}^2}{2}\right]_6^{10}$	M1
	$= \frac{16}{25}$	A1
		(2)
(c)	$\mathrm{E}(T) = \int_0^{10} kt^2 dt = \left[\frac{\mathrm{kt}^3}{3}\right]_0^{10}$	M1
	$= 6\frac{2}{3}$	A1
	$\mathrm{Var}\ (\mathrm{T}) = \int_0^{10} kt^3 dt - \left(6\frac{2}{3}\right)^2 = \left[\frac{kt^4}{4}\right]_0^{10}; -\left(6\frac{2}{3}\right)^2$	M1;M1dep
	$= 50 - \left(6\frac{2}{3}\right)^2$	
	$= 5\frac{5}{9}$	A1
		(5)
(d)	10	B1
		(1)
(e)		B1
		(1)

Question Number	Scheme	Marks
5 (a)	X represents the number of defective components.	
	P ($X = 1$) $= (0.99)^9 (0.01) \times 10$ $= 0.0914$	M1A1
		(2)
(b)	P ($X \geq 2$) = 1 – P($X \leq 1$) = 1 – $(p)^{10}$ – (a) = 0.0043	M1 A1√ A1
		(3)
(c)	$X \sim$ Po(2.5)	B1B1
	P(1 $\leq X \leq$ 4) = P($X \leq$ 4) - P($X = 0$)	M1
	= 0.8912 – 0.0821	
	= 0.809	A1
		(4)
	Normal distribution used. B1for mean only	
	Special case for parts a and b If they use 0.1 do not treat as misread as it makes it easier. (a) M1 A0 if they have 0.3874 (b) M1 A1ft A0 they will get 0.2639 (c) Could get B1 B0 M1 A0	
	For any other values of p which are in the table do not use misread. Check using the tables. They could get (a) M1 A0 (b) M1 A1ft A0 (c) B1 B0 M1 A0	

Question Number	Scheme	Marks
6 (a)(i)	$H_0 : \lambda = 7$ $H_1 : \lambda > 7$	B1
	X = number of visits. $X \sim Po(7)$	B1
	$P(X \geq 10) = 1 - P(X \leq 9)$ $= 0.1695$ $1 - P(X \leq 10) = 0.0985$ $1 - P(X \leq 9) = 0.1695$	M1
	CR $X \geq 11$	A1
	$0.1695 > 0.10$, CR $X \geq 11$	
	Not significant or it is not in the critical region or do not reject H_0	M1
	The rate of visits on a Saturday is not greater/ is unchanged	A1 no ft
(ii)	$X = 11$	B1 (7)
(b)	(The visits occur) randomly/ independently or singly or constant rate	B1 (1)
(c)	[$H_0 : \lambda = 7$ $H_1 : \lambda > 7$ (or $H_0 : \lambda = 14$ $H_1 : \lambda > 14$)]	
	$X \sim$N;(14,14)	B1;B1
	$P(X \geq 20) = P\left(z \geq \frac{19.5 - 14}{\sqrt{14}}\right)$ +/- 0.5, stand	M1 M1
	$= P(z \geq 1.47)$	
	$= 0.0708$ or $z = 1.2816$	A1dep both M
	$0.0708 < 0.10$ therefore significant. The rate of visits is greater on a Saturday	A1dep 2nd M (6)

Question Number	Scheme	Marks
7 (a)	$\mathrm{F}(x_0) = \int_1^x -\frac{2}{9}x + \frac{8}{9}\,\mathrm{dx} = \left[-\frac{1}{9}x^2 + \frac{8}{9}x\right]_1^x$	M1A1
	$= \left[-\frac{1}{9}x^2 + \frac{8}{9}x\right] - \left[-\frac{1}{9} + \frac{8}{9}\right]$	
	$= -\frac{1}{9}x^2 + \frac{8}{9}x - \frac{7}{9}$	A1
		(3)
(b)	$\mathrm{F}(x) = \begin{cases} 0 & x < 1 \\ -\frac{1}{9}x^2 + \frac{8}{9}x - \frac{7}{9} & 1 \le x \le 4 \\ 1 & x > 4 \end{cases}$	B1B1√
		(2)
(c)	$\mathrm{F}(x) = 0.75$; or $\mathrm{F}(2.5) = -\frac{1}{9} \times 2.5^2 + \frac{8}{9} \times 2.5 - \frac{7}{9}$	M1;
	$-\frac{1}{9}x^2 + \frac{8}{9}x - \frac{7}{9} = 0.75$	
	$4x^2 - 32^x + 55 = 0$	
	$-x^2 + 8x - 13.75 = 0$	
	$x = 2.5$ $= 0.75$ cso	A1
	and $\mathrm{F}(x) = 0.25$	
	$-\frac{1}{9}x^2 + \frac{8}{9}x - \frac{7}{9} = 0.25$	M1
	$-x^2 + 8x - 7 = 2.25$	
	$-x^2 + 8x - 9.25 = 0$ quadratic 3 terms =0	M1 dep
	$x = \dfrac{-8 \pm \sqrt{8^2 - 4 \times -1 \times -9.25}}{2 \times -1}$	M1 dep
		A1
	$x = 1.40$	
		(6)
(d)	$Q_3 - Q_2 > Q_2 - Q_1$ Or mode = 1 and mode < median Or mean = 2 and median < mode Sketch of pdf here or be referred to if in a different part of the question Box plot with Q_1, Q_2, Q_3 values marked on	M1
	Positive skew	A1
		(2)