Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Advanced

Tuesday 15 January 2008 – Morning

Time: 1 hour 30 minutes

Materials required for examination
Mathematical Formulae (Green)

Items included with question papers
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 8 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.

N29278A

1. (*a*) Explain what you understand by a census.

(1)

Each cooker produced at GT Engineering is stamped with a unique serial number. GT Engineering produces cookers in batches of 2000. Before selling them, they test a random sample of 5 to see what electric current overload they will take before breaking down.

(*b*) Give one reason, other than to save time and cost, why a sample is taken rather than a census.

(1)

(*c*) Suggest a suitable sampling frame from which to obtain this sample.

(1)

(*d*) Identify the sampling units.

(1)

---

2. The probability of a bolt being faulty is 0.3. Find the probability that in a random sample of 20 bolts there are

(*a*) exactly 2 faulty bolts,

(2)

(*b*) more than 3 faulty bolts.

(2)

These bolts are sold in bags of 20. John buys 10 bags.

(*c*) Find the probability that exactly 6 of these bags contain more than 3 faulty bolts.

(3)

---

3. (*a*) State two conditions under which a Poisson distribution is a suitable model to use in statistical work.

(2)

The number of cars passing an observation point in a 10 minute interval is modelled by a Poisson distribution with mean 1.

(*b*) Find the probability that in a randomly chosen 60 minute period there will be

(i) exactly 4 cars passing the observation point,

(ii) at least 5 cars passing the observation point.

(5)

The number of other vehicles, other than cars, passing the observation point in a 60 minute interval is modelled by a Poisson distribution with mean 12.

(*c*) Find the probability that exactly 1 vehicle, of any type, passes the observation point in a 10 minute period.

(4)

---

4. The continuous random variable Y has cumulative distribution function F(y) given by

$$\mathrm{F}(y) = \begin{cases} 0 & y < 1 \\ k(y^4 + y^2 - 2) & 1 \le y \le 2 \\ 1 & y > 2 \end{cases}$$

(a) Show that $k = \frac{1}{18}$.

(2)

(b) Find P($Y > 1.5$).

(2)

(c) Specify fully the probability density function f(y).

(3)

---

5. Dhriti grows tomatoes. Over a period of time, she has found that there is a probability 0.3 of a ripe tomato having a diameter greater than 4 cm. She decides to try a new fertiliser. In a random sample of 40 ripe tomatoes, 18 have a diameter greater than 4 cm. Dhriti claims that the new fertiliser has increased the probability of a ripe tomato being greater than 4 cm in diameter.

Test Dhriti's claim at the 5% level of significance. State your hypotheses clearly.

(7)

---

6. The probability that a sunflower plant grows over 1.5 metres high is 0.25. A random sample of 40 sunflower plants is taken and each sunflower plant is measured and its height recorded.

(a) Find the probability that the number of sunflower plants over 1.5 m high is between 8 and 13 (inclusive) using

(i) a Poisson approximation,

(ii) a Normal approximation.

(10)

(b) Write down which of the approximations used in part (a) is the most accurate estimate of the probability. You must give a reason for your answer.

(2)

---

7. (*a*) Explain what you understand by

(i) a hypothesis test,

(ii) a critical region.

(3)

During term time, incoming calls to a school are thought to occur at a rate of 0.45 per minute. To test this, the number of calls during a random 20 minute interval, is recorded.

(*b*) Find the critical region for a two-tailed test of the hypothesis that the number of incoming calls occurs at a rate of 0.45 per 1 minute interval. The probability in each tail should be as close to 2.5% as possible.

(5)

(*c*) Write down the actual significance level of the above test.

(1)

In the school holidays, 1 call occurs in a 10 minute interval.

(*d*) Test, at the 5% level of significance, whether or not there is evidence that the rate of incoming calls is less during the school holidays than in term time.

(5)

---

8. The continuous random variable X has probability density function f(x) given by

$$\text{f}(x) = \begin{cases} 2(x-2) & 2 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

(*a*) Sketch f(x) for all values of x.

(3)

(*b*) Write down the mode of X.

(1)

Find

(*c*) E(X),

(3)

(*d*) the median of X.

(4)

(*e*) Comment on the skewness of this distribution. Give a reason for your answer.

(2)

TOTAL FOR PAPER: 75 MARKS

END

January 2008
Statistics S2
Mark Scheme

Question Number	Scheme	Marks
1. (a)	A census is when <u>every member</u> of the <u>population</u> is investigated.	B1
(b)	There would be no cookers left to sell.	B1
(c)	A list of the unique identification numbers of the cookers.	B1
(d)	A cooker	B1
		(4)
Notes		
1. (a)	**B1** Need one word from each group (1) <u>Every member /all items / entire /oe</u> (2) <u>population/collection of individuals/sampling frame/oe</u> enumerating the population on its own gets B0	
(b)	**B1** Idea of Tests to destruction. Do not accept cheap or quick	
(c)	**B1** Idea of list/ register/database of cookers/serial numbers	
(d)	**B1** cooker(s) / serial number(s) The sample of 5 cookers or every 400^{th} cooker gets B1	

2 (a)	Let X be the random variable the number of faulty bolts	M1
	$P(X \leq 2) - P(X \leq 1) = 0.0355 - 0.0076$ or $(0.3)^2(0.7)^{18}\frac{20!}{18!2!}$ $= 0.0279$ $= 0.0278$	A1 (2)
(b)	$1 - P(X \leq 3) = 1 - 0.1071$ $= 0.8929$	M1 A1 (2)
	or $1-(0.3)^3(0.7)^{17}\frac{20!}{17!3!}-(0.3)^2(0.7)^{18}\frac{20!}{18!2!}-(0.3)(0.7.)^{19}\frac{20!}{19!1!}-(0.7)^{20}$	M1A1√A1
(c)	$\frac{10!}{4!6!}(0.8929)^6(0.1071)^4 = 0.0140.$	(3)

Notes:		
2. (a)	**M1** Either attempting to use P ($X \leq 2$) – P ($X \leq 1$) or attempt to use binomial and find p($X = 2$). Must have $(p)^2(1-p)^{18}\frac{20!}{18!2!}$, with a value of p **A1** awrt 0.0278 or 0.0279.	
(b)	**M1** Attempting to find $1 - P(X \leq 3)$ **A1** awrt 0.893	
(c)	**M1** for $k\ (p)^6(1-p)^4$. They may use any value for p and k can be any number or ${}^nC_6p^6(1-p)^{n-6}$ **A1**√ $\frac{10!}{4!6!}(their\ part\ b)^6(1-their\ part\ b)^4$ may write ${}^{10}C_6$or ${}^{10}C_4$ **A1** awrt 0.014	B1 B1 (2)

3. (a)	Events occur at a constant rate. any two of the 3 Events occur independently or randomly. Events occur singly.	B1
		M1
(b)	Let X be the random variable the number of cars passing the observation point.	A1
	Po(6)	M1
(i)	$P(X \leq 4) - P(X \leq 3) = 0.2851 - 0.1512$ or $\frac{e^{-6}6^4}{4!}$	A1
	$= 0.1339$	(5)
	$1 - P(X \leq 4) = 1 - 0.2851$ or $1 - e^{-6}\left(\frac{6^4}{4!} + \frac{6^3}{3!} + \frac{6^2}{2!} + \frac{6}{1!} + 1\right)$	B1
(ii)	$= 0.7149$	M1 A1
		A1 (4)
(c)	P (0 car and 1 others) + P (1 cars and 0 other) $= e^{-1} \times 2e^{-2} + 1e^{-1} \times e^{-2}$ $= 0.3679 \times 0.2707 + 0.3674 \times 0.1353$ $= 0.0996 + 0.0498$ $= 0.149$ alternative $P_o(1+2) = P_o(3)$ B1 $P(X=1) = 3e^{-3}$ M1 A1 $= 0.149$ A1	

Notes
3(a)

B1 B1 Need the word events at least once.
Independently and randomly are the same reason.
Award the first B1 if they only gain 1 mark
Special case. If they have 2 of the 3 lines without the word events they get B0 B1

B1 Using Po(6) in (i) or (ii)

(b) (i) **M1** Attempting to find $P(X \leq 4) - P(X \leq 3)$ or $\frac{e^{-\lambda}\lambda^4}{4!}$

A1 awrt 0.134

M1 Attempting to find $1 - P(X \leq 4)$

(ii) **A1** awrt 0.715

(c) **B1** Attempting to find both possibilities. May be implied by doing $e^{-\lambda_1} \times \lambda_2 e^{-\lambda_2} + e^{-\lambda_2} \times \lambda_1 e^{-\lambda_1}$ any values of λ_1 and λ_2

M1 finding one pair of form $e^{-\lambda_1} \times \lambda_2 e^{-\lambda_2}$ any values of λ_1 and λ_2

A1 one pair correct

A1 awrt 0.149

Alternative.
B1 for Po(3)
M1 for attempting to find $P(X=1)$ with Po(3)
A1 $3e^{-3}$
A1 awrt 0.149

4. (a)	$K(2^4 + 2^2 - 2) = 1$ $K = 1/18$	M1 A1 (2)
(b)	$1 - F(1.5) = 1 - \frac{1}{18}(1.5^4 + 1.5^2 - 2)$ $= 0.705$ or $\frac{203}{288}$	M1 A1 (2)
(c)	$f(y) = \begin{cases} \frac{1}{9}(2y^3 + y) & 1 \le y \le 2 \\ 0 & \textit{otherwise} \end{cases}$	M1 A1 B1 (3)
Notes 4. (a)	**M1** putting F(*2*) = 1 or F(*2*) – F(1) = 1 **A1** cso. Must show substituting $y = 2$ and the 1/18	
(b)	**M1** either attempting to find 1 – F(1.5) may write and use F(2) – F (1.5) **A1** awrt 0.705	
(c)	**M1** attempting to differentiate. Must see either a $y^n \rightarrow y^{n-1}$ at least once **A1** for getting $\frac{1}{9}(2y^3 + y)$ o.e and $1 \le y \le 2$ allow $1 < y < 2$ **B1** for the 0 *otherwise.* Allow 0 for y <1 and 0 for y >2 Allow them to use any letter	

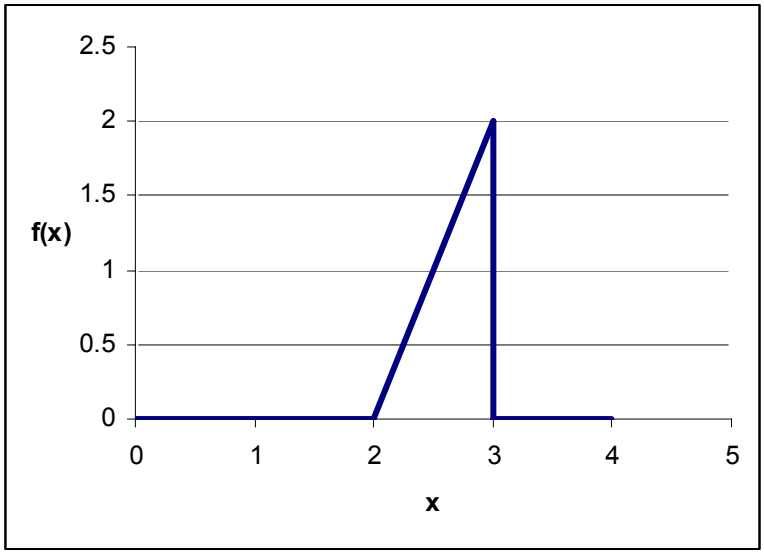
Question	Scheme	Marks
5	$H_0 : p = 0.3$; $H_1 : p > 0.3$ Let X represent the number of tomatoes greater than 4 cm : $X \sim B(40, 0.3)$ $P(X \geq 18) = 1 - P(X \leq 17)$ $\quad$ $P(X \geq 18)$ $1 - P(X \leq 17) = 0.0320$ $P(X \geq 17) = 1 - P(X \leq 16) = 0.0633$ $= 0.0320$ $\quad$ CR $X \geq 18$ $0.0320 < 0.05$ $\quad$ $18 \geq 18$ or 18 in the critical region no evidence to Reject H_0 or it is significant New fertiliser has <u>increased</u> the probability of a <u>tomato</u> being greater than 4 cm **Or** Dhriti's claim is true	B1 B1 B1 M1 A1 M1 B1d cao (7)
5	**B1** for correct H_0 . must use p or pi **B1** for correct H_1 must use p and be one tail. **B1** using B(40, 0.3). This may be implied by their calculation **M1** attempt to find $1 - P(X \leq 17)$ or get a correct probability. For CR method must attempt to find $P(X \geq 18)$ or give the correct critical region **A1** awrt 0.032 or correct CR. **M1** correct statement based on their probability , H_1 and 0.05 or a correct contextualised statement that implies that. **B1** this is not a follow through .conclusion in context. Must use the words increased, tomato and some reference to size or diameter. This is dependent on them getting the previous M1 If they do a **<u>two tail test</u>** they may get B1 B0 B1 M1 A1 M1 B0 For the second M1 they must have accept Ho or it is not significant or a correct contextualised statement that implies that. .	

6a (i)	Let X represent the number of sunflower plants more than 1.5m high	B1
	$X \sim$ Po(10) $\mu=10$	M1
	$P(8 \leq X \leq 13) = P(X \leq 13) - P(X \leq 7)$	
	$= 0.8645 - 0.2202$	A1
	$= 0.6443$ awrt 0.644	B1
ii)	$X \sim N(10,7.5)$	M1 M1
	$P(7.5 \leq X \leq 13.5) = P\left(\frac{7.5-10}{\sqrt{7.5}} \leq X \leq \frac{13.5-10}{\sqrt{7.5}}\right)$	A1 A1
	$= P(-0.913 \leq X \leq 1.278)$	M1
	$= 0.8997 - (1 - 0.8186)$	A1 (10)
	$= 0.7183$ awrt 0.718 or 0.719	B1
b)	Normal approx /not Poisson since (n is large) and p close to half. **or** (np = 10 npq = 7.5) mean ≠ variance **or** np (= 10) and nq (= 30) both >5. or exact binomial = 0.7148	B1dep (2)
6a (i)	**B1** mean = 10 May be implied in (i) or (ii)	
	M1 Attempting to find $P(X \leq 13) - P(X \leq 7)$	
	A1 awrt 0.644	
ii)	**B1** $\sigma^2 = 7.5$ May be implied by being correct in standardised formula	
	M1 using 7.5 or 8.5 or 12.5 or 13.5.	
	M1 standardising using 7.5 or 8 or 8.5 **or** 12.5 or13 or 13.5 and their mean and standard deviation.	
	A1 award for either $\frac{7.5-10}{\sqrt{7.5}}$ or awrt -0.91	
	A1 award for either $\frac{13.5-10}{\sqrt{7.5}}$ or awrt 1.28	
	M1 Finding the correct area. Following on from their 7.5 and 13.5. Need to do a Prob >0.5 – prob <0.5 or prob <0.5 + prob< 0.5	

b)	**A1** awrt 0.718 or 0.719 only. Dependent on them getting all three method marks. No working but correct answer will gain all the marks first **B1** normal second **B1** p close to half, or mean ≠ variance or np and nq both > 5.They may use a number bigger than 5 or they may work out the exact value 0.7148 using the binomial distribution. **Do not allow np> 5 and npq>5**	

7 ai)	A hypothesis test is a mathematical procedure to examine a value of a population parameter proposed by the null hypothesis compared with an alternative hypothesis.	B1
ii)	The critical region is the range of values **or** a test statistic or region where the test is significant that would lead to the rejection of H_0.	B1g B1h (3)
(b)	Let X represent the number of incoming calls : $X \sim Po(9)$	B1
	From table $P(X \geq 16) = 0.0220$	M1 A1
	$P(x \leq 3) = 0.0212$	A1
	Critical region ($x \leq 3$ or $x \geq 16$)	B1 (5)
(c)	Significance level = 0.0220 + 0.0212 = 0.0432 or4.32%	B1 (1)
(d)	$H_o : \lambda = 0.45$; $H_1 : \lambda < 0.45$ (accept : $H_o : \lambda = 4.5$; $H_1 : \lambda < 4.5$)	B1
	Using $X \sim Po(4.5)$	M1
	$P(X \leq 1) = 0.0611$ CR $X \leq 0$ awrt 0.0611	A1
	0.0611 > 0.05. $1 \geq 0$ or 1not in the critical region	M1
	There is evidence to Accept H_0 or it is not significant	B1cao
	There is no evidence that there are less calls during school holidays.	(5)
Notes 7 ai)	B1 Method for deciding between 2 hypothesis.	
ii)	B1 range of values. This may be implied by other words. Not region on its own B1 which lead you to reject H_0 Give the first B1 if only one mark awarded.	
(b)	B1 using $P_o(9)$	
	M1 attempting to find $P(X \geq 16)$ or $P(x \leq 3)$	
	A1 0.0220 or $P(X \geq 16)$	

	A1 0.0212 or P(x ≤ 3) These 3 marks may be gained by seeing the numbers in part c B1 correct critical region A completely correct critical region will get all 5 marks. Half of the correct critical region eg x ≤ 3 or x ≥ 17 say would get B1 M1 A0 A1 B0 if the M1 A1 A1 not already awarded.	
(c)	B1 cao awrt 0.0432	
(d)	B1 may use λ or μ. Needs both H_0 and H_1 M1 using $P_o(4.5)$ A1 correct probability or CR only M1 correct statement based on their probability , H_1 and 0.05 or a correct contextualised statement that implies that. **B1** this is not a follow through .Conclusion in context. Must see the word **calls** in conclusion If they get the correct CR with no evidence of using $P_o(4.5)$ they will get M0 A0 SC If they get the critical region $X \leq 1$ they score M1 for rejecting H_0 and B1 for concluding the rate of calls in the holiday is lower.	

8. a)	 Max height of 2 labelled and goes through(2,0) shape must be between 2 and 3 and no other lines drawn (accept patios drawn) correct shape	B1 B1 B1 (3)
		B1 (1)
b)	3	
c)	$\int_2^3 2x(x-2)\,dx = \left[\frac{2x^3}{3} - 2x^2\right]_2^3$	M1A1
	$= 2\frac{2}{3}$	A1 (3)
d)	$\int_2^m 2(x-2)\,dx = 0.5$	M1
	$\left[x^2 - 4x\right]_2^m = 0.5$	
	$m^2 - 4m + 4 = 0.5$	A1
	$m^2 - 4m + 3.5 = 0$	
	$m = \frac{4 \pm \sqrt{2}}{2}$	M1
	m =2.71	A1 (4)
e)	Negative skew. mean < median < mode .	B1 B1dep (2)

Notes 8.		
(a)	**B1** the graph must have a maximum of 2 which must be labelled **B1** the line must be between 2 and 3 with not other line drawn except patios. They can get this mark even if the patio cannot be seen. **B1** the line must be straight and the right shape.	
(b)	**B1** Only accept 3	
(c)	**M1** attempt to find $\int x\mathrm{f}(x)\mathrm{d}x$ for attempt we need to see $x^{\mathrm{n}} \rightarrow x^{\mathrm{n+1}}$. ignore limits **A1** correct integration ignore limits **A1** accept $2\frac{2}{3}$ or awrt 2.67 or $2.\dot{6}$	
(d)	**M1** using $\int \mathrm{f}(x)\mathrm{d}x = 0.5$ **A1** $m^2 - 4m + 4 = 0.5$ oe **M1** attempting to solve quadratic. **A1** awrt 2.71 or $\frac{4+\sqrt{2}}{2}$ or $2+\frac{\sqrt{2}}{2}$ oe	
(e)	First **B1** for negative Second **B1** for mean < median< mode. Need all 3 or may explain using diagram.	