Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Advanced Level

Monday 1 June 2009 – Morning

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Orange or 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 not gain full credit.

M34280A

1. A bag contains a large number of counters of which 15% are coloured red. A random sample of 30 counters is selected and the number of red counters is recorded.

(*a*) Find the probability of no more than 6 red counters in this sample.

(2)

A second random sample of 30 counters is selected and the number of red counters is recorded.

(*b*) Using a Poisson approximation, estimate the probability that the total number of red counters in the combined sample of size 60 is less than 13.

(3)

---

2. An effect of a certain disease is that a small number of the red blood cells are deformed. Emily has this disease and the deformed blood cells occur randomly at a rate of 2.5 per ml of her blood. Following a course of treatment, a random sample of 2 ml of Emily's blood is found to contain only 1 deformed red blood cell.

Stating your hypotheses clearly and using a 5% level of significance, test whether or not there has been a decrease in the number of deformed red blood cells in Emily's blood.

(6)

---

3. A random sample $X_1, X_2, \ldots X_n$ is taken from a population with unknown mean μ and unknown variance σ^2. A statistic Y is based on this sample.

(*a*) Explain what you understand by the statistic Y.

(2)

(*b*) Explain what you understand by the sampling distribution of Y.

(1)

(*c*) State, giving a reason which of the following is **not** a statistic based on this sample.

(i) $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n}$ (ii) $\displaystyle\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$ (iii) $\displaystyle\sum_{i=1}^{n} X_i^{\,2}$

(2)

---

4. Past records suggest that 30% of customers who buy baked beans from a large supermarket buy them in single tins. A new manager questions whether or not there has been a change in the proportion of customers who buy baked beans in single tins. A random sample of 20 customers who had bought baked beans was taken.

(*a*) Using a 10% level of significance, find the critical region for a two-tailed test to answer the manager's question. You should state the probability of rejection in each tail which should be less than 0.05.

(5)

(*b*) Write down the actual significance level of a test based on your critical region from part (*a*).

(1)

The manager found that 11 customers from the sample of 20 had bought baked beans in single tins.

(*c*) Comment on this finding in the light of your critical region found in part (*a*).

(2)

---

5. An administrator makes errors in her typing randomly at a rate of 3 errors every 1000 words.

(*a*) In a document of 2000 words find the probability that the administrator makes 4 or more errors.

(3)

The administrator is given an 8000 word report to type and she is told that the report will only be accepted if there are 20 or fewer errors.

(*b*) Use a suitable approximation to calculate the probability that the report is accepted.

(7)

---

6. The three independent random variables A, B and C each has a continuous uniform distribution over the interval [0, 5].

(*a*) Find $P(A > 3)$.

(1)

(*b*) Find the probability that A, B and C are all greater than 3.

(2)

The random variable Y represents the maximum value of A, B and C.

The cumulative distribution function of Y is

$$F(y) = \begin{cases} 0, & y < 0 \\ \dfrac{y^3}{125}, & 0 \leq y \leq 5 \\ 1, & y > 5 \end{cases}$$

(*c*) Find the probability density function of Y.

(2)

(*d*) Sketch the probability density function of Y.

(2)

(*e*) Write down the mode of Y.

(1)

(*f*) Find $E(Y)$.

(3)

(*g*) Find $P(Y > 3)$.

(2)

7.

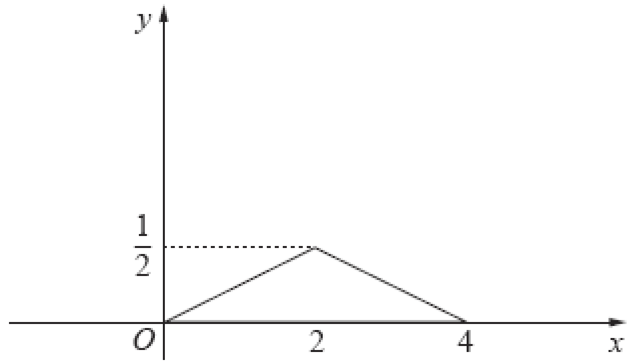


Figure 1

Figure 1 shows a sketch of the probability density function f(x) of the random variable X. The part of the sketch from $x = 0$ to $x = 4$ consists of an isosceles triangle with maximum at (2, 0.5).

(a) Write down E(X).

(1)

The probability density function f(x) can be written in the following form.

$$f(x) = \begin{cases} ax & 0 \leq x < 2 \\ b - ax & 2 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

(b) Find the values of the constants a and b.

(2)

(c) Show that σ, the standard deviation of X, is 0.816 to 3 decimal places.

(7)

(d) Find the lower quartile of X.

(3)

(e) State, giving a reason, whether P($2 - \sigma < X < 2 + \sigma$) is more or less than 0.5

(2)

8. A cloth manufacturer knows that faults occur randomly in the production process at a rate of 2 every 15 metres.

(*a*) Find the probability of exactly 4 faults in a 15 metre length of cloth.

(2)

(*b*) Find the probability of more than 10 faults in 60 metres of cloth.

(3)

A retailer buys a large amount of this cloth and sells it in pieces of length x metres. He chooses x so that the probability of no faults in a piece is 0.80.

(*c*) Write down an equation for x and show that $x = 1.7$ to 2 significant figures.

(4)

The retailer sells 1200 of these pieces of cloth. He makes a profit of 60p on each piece of cloth that does not contain a fault but a loss of £1.50 on any pieces that do contain faults.

(*d*) Find the retailer's expected profit.

(4)

TOTAL FOR PAPER: 75 MARKS

END

June 2009
6684 Statistics S2
Mark Scheme

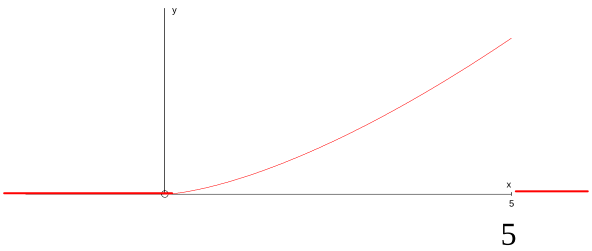
Question Number	Scheme	Marks
Q1 (a)	$[X \sim B(30, 0.15)]$	
	$P(X \le 6)$, = 0.8474 awrt 0.847	M1, A1 (2)
(b)	$Y \sim B(60, 0.15) \approx Po(9)$ for using Po(9)	B1
	$P(Y \le 12)$, = 0.8758 awrt 0.876	M1, A1 (3)
		[5]
	[N.B. normal approximation gives 0.897, exact binomial gives 0.894]	
(a) (b)	M1 for a correct probability statement $P(X \le 6)$ or $P(X < 7)$ or P (X =0) + P (X = 1)+ P (X = 2)+ P (X =4) + P (X =5) + P (X =6). (may be implied by long calculation) Correct answer gets M1 A1. allow 84.74% B1 may be implied by using Po(9). Common incorrect answer which implies this is 0.9261 M1 for a correct probability statement $P(X \le 12)$ or $P(X < 13)$ or P (X =0)+ P (X = 1)+…+ P (X = 12) (may be implied by long calculation) and attempt to evaluate this probability using their Poisson distribution. Condone P ($X \le 13$) = 0.8758 for B1 M1 A1 Correct answer gets B1 M1 A1 Use of normal or exact binomial get B0 M0 A0	

Question Number	Scheme	Marks
Q2	H_0: $\lambda = 2.5$ (or $\lambda = 5$) H1: $\lambda < 2.5$ (or $\lambda < 5$) λ or μ	B1B1
	$X \sim \text{Po}(5)$	M1
	$P(X \le 1) = 0.0404$ or CR $X \le 1$	A1
	[0.0404<0.05] this is significant or reject H_0 or it is in the critical region	M1
	There is evidence of a decrease in the (mean) number/rate of deformed blood cells	A1 (6) **[6]**
	1st B1 for H_0 must use lambda or mu; 5 or 2.5. 2nd B1 for H_1 must use lambda or mu; 5 or 2.5 1st M1 for use of Po(5) may be implied by probability(must be used not just seen) eg. P (X = 1) = 0.0404 - … would score M1 A0 1st A1 for 0.0404 seen or correct CR 2nd M1 for a correct statement (this may be contextual) comparing their probability and 0.05 (or comparing 1 with their critical region). Do not allow conflicting statements. 2nd A1 is not a follow through. Need the word decrease, number or rate and deformed blood cells for contextual mark. If they have used $\neq$ in H_1 they could get B1 B0 M1 A1 M1A0 mark as above except they gain the 1st A1 for $P(X \le 1) = 0.0404$ or CR $X \le 0$ 2nd M1 for a correct statement (this may be contextual) comparing their probability and 0.025 (or comparing 1 with their critical region) They may compare with 0.95 (one tail method) or 0.975 (one tail method) Probability is 0.9596.	

Question Number	Scheme	Marks
Q3 (a)	*A statistic* is a function of $X_1, X_2, \ldots X_n$ that does not contain any unknown parameters	B1 B1 (2)
(b)	The probability distribution of Y or the distribution of all possible values of Y (o.e.)	B1 (1)
(c)	Identify (ii) as not a statistic Since it contains unknown parameters μ and σ.	B1 dB1 (2)
		[5]
(a)	Examples of other acceptable wording: B1 e.g. is a function of the sample or the data / is a quantity calculated from the sample or the data / is a random variable calculated from the sample or the data B1 e.g. does not contain any unknown parameters/quantities contains only known parameters/quantities only contains values of the sample Y is a function of $X_1, X_2, \ldots X_n$ that does not contain any unknown parameters B1B1 is a function of the values of a sample with no unknowns B1B1 is a function of the sample values B1B0 is a function of all the data values B1B0 A random variable calculated from the sample B1B0 A random variable consisting of any function B0B0 A function of a value of the sample B1B0 A function of the sample which contains no other values/ parameters B1B0	
(b)	Examples of other acceptable wording All possible values of the statistic together with their associated probabilities	
(c)	1st B1 for selecting only (ii) 2nd B1 for a reason. This is dependent upon the first B1. Need to mention at least one of mu (mean) or sigma (standard deviation or variance) or unknown parameters. Examples since it contains mu B1 since it contains sigma B1 since it contains unknown parameters/quantities B1 since it contains unknowns B0	

Question Number	Scheme	Marks
Q4 (a)	X~ B(20, 0.3) $\mathrm{P}(X \le 2) = 0.0355$ $\mathrm{P}(X \le 9) = 0.9520$ so $\mathrm{P}(X \ge 10) = 0.0480$ Therefore the critical region is $\{X \le 2\} \cup \{X \ge 10\}$	M1 A1 A1 A1A1 (5)
(b)	0.0355 + 0.0480 = 0.0835 awrt (0.083 or 0.084)	B1 (1)
(c)	11 is in the critical region there is evidence of a <u>change/ increase</u> in the <u>proportion/number</u> of <u>customers buying single tins</u>	B1ft B1ft (2) **[8]**
(a)	M1 for B(20,0.3) seen or used 1st A1 for 0.0355 2nd A1 for 0.048 3rd A1 for $(X) \le 2$ or $(X) < 3$ or [0,2] They get **A0** if they write $\mathrm{P}(X \le 2/ X < 3)$ 4th A1 $(X) \ge 10$ or $(X) > 9$ or [10,20] They get **A0** if they write $\mathrm{P}(X \ge 10/ X > 9)$ **10** $\le X \le 2$ etc is accepted To describe the critical regions they can use any letter or no letter at all. It does not have to be X.	
(b)	B1 correct answer only	
(c)	1st B1 for a correct statement about 11 and their critical region. 2nd B1 for a correct comment in context consistent with their CR and the value 11 Alternative solution 1st B0 $P(X \ge 11) = 1 - 0.9829 = 0.0171$ since no comment about the critical region 2nd B1 a correct contextual statement.	

Question Number	Scheme	Marks
Q5 (a)	X = the number of errors in 2000 words so $X \sim \text{Po}(6)$	B1
	$P(X \geq 4) = 1 - P(X \leq 3)$	M1
	$= 1 - 0.1512 \quad = 0.8488$ awrt 0.849	A1 (3)
(b)	Y = the number of errors in 8000 words. $Y \sim \text{Po}(24)$ so use a <u>Normal</u> approx	M1
	$Y \approx\sim \text{N}(24, \sqrt{24}^2)$	A1
	Require $P(Y \leq 20) = P\left(Z < \frac{20.5 - 24}{\sqrt{24}}\right)$	M1 M1
	$= P(Z < -0.714\ldots)$	A1
	$= 1 - 0.7611$	M1
	$= 0.2389$ awrt (0.237~0.239)	A1 (7)
	[N.B. Exact Po gives 0.242 and no $\pm$ 0.5 gives 0.207]	**[10]**
(a)	B1 for seeing or using Po(6) M1 for $1 - P(X \leq 3)$ or $1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$ A1 awrt 0.849	
SC	If B(2000, 0.003) is used and leads to awrt 0.849 allow B0 M1 A1 If no distribution indicated awrt 0.8488 scores B1M1A1 but any other awrt 0.849 scores B0M1A1	
(b)	1st M1 for identifying the normal approximation 1st A1 for [mean = 24] **and** [sd = $\sqrt{24}$ or var = 24] These first two marks may be given if the following are seen in the standardisation formula : 24 $\sqrt{24}$ or awrt 4.90 2nd M1 for attempting a continuity correction (20/ 28 $\pm$ 0.5 is acceptable) 3rd M1 for standardising using their mean and their standard deviation. 2nd A1 correct z value awrt $\pm$0.71 **or** this may be awarded if see $\frac{20.5 - 24}{\sqrt{24}}$ or $\frac{27.5 - 24}{\sqrt{24}}$ 4th M1 for 1 - a probability from tables (must have an answer of < 0.5) 3rd A1 answer awrt 3 sig fig in range 0.237 – 0.239	

Question Number	Scheme	Marks
Q6 (a)	P(A >3) = $\frac{2}{5}$ = 0.4	B1 (1)
(b)	$(0.4)^3$,= 0.064 or $\frac{8}{125}$	M1, A1 (2)
(c)	$\text{f}(y) = \frac{\text{d}}{\text{d}y}(\text{F}(y)) = \begin{cases} \frac{3y^2}{125} & 0 \le y \le 5 \\ 0 & otherwise \end{cases}$	M1A1 (2)
(d)	 Shape of curve and start at (0,0) Point (5, 0) labelled and curve between 0 and 5 and pdf $\geq$ 0	B1 B1 (2)
(e)	Mode = 5	B1 (1)
(f)	$\text{E}(Y) = \int_0^5 \left(\frac{3y^3}{125}\right) \text{d}y = \left[\frac{3y^4}{500}\right]_0^5 = \frac{15}{4}$ or 3.75	M1M1A1 (3)
(g)	$\text{P}(Y > 3) = \begin{cases} \int_3^5 \frac{3y^2}{125} \text{d}y \\ \text{or } 1 - \text{F}(3) \end{cases} = 1 - \frac{27}{125} = \frac{98}{125} = 0.784$	M1A1 (2) **[13]**
(a)	B1 correct answer only(cao). Do not ignore subsequent working	
(b)	M1 for cubing their answer to part (a) A1 cao	
(c)	M1 for attempt to differentiate the cdf. They must decrease the power by 1 A1 fully correct answer including 0 otherwise. Condone < signs	
(d)	B1 for shape. Must curve the correct way and start at (0,0). No need for y = 0 (patios) lines B1 for point (5,0) labelled and pdf only existing between 0 and 5, may have y=0 (patios) for other values	
(e)	B1 cao	
(f)	1st M1 for attempt to integrate their $y\text{f}(y)$ $y^n \to y^{n+1}$. 2nd M1 for attempt to use correct limits A1 cao	
(g)	M1 for attempt to find P(Y > 3). e.g. writing $\int_3^5 their\ f(y)$ must have correct limits or writing 1 – F(3)	

Question Number	Scheme	Marks
Q7 (a)	E(X) = 2 (by symmetry)	B1 (1)
(b)	$0 \le x < 2$, gradient = $\frac{\frac{1}{2}}{2} = \frac{1}{4}$ and equation is $y = \frac{1}{4}x$ so $a = \frac{1}{4}$	B1
	$b - \frac{1}{4}x$ passes through (4, 0) so $b = 1$	B1 (2)
(c)	$\text{E}(X^2) = \int_0^2 \left(\frac{1}{4}x^3\right)\,\text{d}x + \int_2^4 \left(x^2 - \frac{1}{4}x^3\right)\,\text{d}x$	M1M1
	$= \left[\frac{x^4}{16}\right]_0^2 + \left[\frac{x^3}{3} - \frac{x^4}{16}\right]_2^4$	A1
	$= 1 + \frac{64-8}{3} - \frac{256-16}{16} = 4\frac{2}{3}$ or $\frac{14}{3}$	M1A1
	Var(X) = $\text{E}(X^2) - [\text{E}(X)]^2 = \frac{14}{3} - 2^2$, $= \frac{2}{3}$ (so $\sigma = \sqrt{\frac{2}{3}} = 0.816$) **(*)**	M1 A1cso (7)
(d)	P($X \le q$) = $\int_0^q \frac{1}{4}x\,\text{d}x = \frac{1}{4}$, $\frac{q^2}{2} = 1$ so $q = \sqrt{2} = 1.414$ awrt 1.41	M1A1,A1 (3)
(e)	2- σ = 1.184 so 2 - σ, 2 + σ is wider than IQR, therefore greater than 0.5	M1,A1 (2) **[15]**

(a) B1 cao

(b) B1 for value of a. B1 for value of b

(c) 1st M1 for attempt at $\int ax^3$ using their a. For attempt they need x^4. Ignore limits.

2nd M1 for attempt at $\int bx^2 - ax^3$ use their a and b. For attempt need to have either x^3 or x^4. Ignore limits

1st A1 correct integration for both parts

3rd M1 for use of the correct limits on each part

2nd A1 for either getting 1 and $3\frac{2}{3}$ or awrt 3.67 somewhere or $4\frac{2}{3}$ or awrt 4.67

4th M1 for use of $\text{E}(X^2) - [\text{E}(X)]^2$ must add both parts for E (X^2) and only have subtracted the mean^2 once. You must see this working

3rd A1 $\sigma = \sqrt{\frac{2}{3}}$ or $\sqrt{0.66667}$ or better with no incorrect working seen.

(d) M1 for attempting to find LQ, integral of either part of f(x) with their 'a' and 'b' = 0.25

Or their F(x) = 0.25 i.e. $\frac{ax^2}{2} = 0.25$ or $bx - \frac{ax^2}{2} + 4a - 2b = 0.25$ with their a and b

If they add both parts of their F(x), then they will get M0.

1st A1 for a correct equation/expression using their 'a'

2nd A1 for $\sqrt{2}$ or awrt 1.41

(e) M1 for a reason based on their quartiles

- Possible reasons are P(2 - σ < X < 2 + σ) = 0.6498 allow awrt 0.65
- 1.184 < LQ(1.414)

A1 for correct answer > 0.5

NB you must check the reason and award the method mark. A correct answer without a correct reason gets M0 A0

Question Number	Scheme	Marks
Q8 (a)	$X \sim \text{Po}(2)$ $\quad$ $P(X = 4) = \dfrac{e^{-2} \times 2^4}{4!} = 0.0902$ $\quad$ awrt 0.09	M1 A1 (2)
(b)	$Y \sim \text{Po}(8)$ $P(Y > 10) = 1 - P(Y \le 10) = 1 - 0.8159 = 0.18411\ldots$ $\quad$ awrt 0.184	B1 M1A1 (3)
(c)	F = no. of faults in a piece of cloth of length x $\quad$ $F \sim \text{Po}(x \times \frac{2}{15})$ $e^{-\frac{2x}{15}} = 0.80$ $e^{-\frac{2}{15}\times 1.65} = 0.8025...,\quad e^{-\frac{2}{15}\times 1.75} = 0.791...$ These values are either side of 0.80 therefore x = 1.7 to 2 sf	 M1A1 M1 A1 (4)
(d)	Expected number with no faults $= 1200 \times 0.8 = 960$ Expected number with some faults $= 1200 \times 0.2 = 240$ So expected profit $= 960 \times 0.60 - 240 \times 1.50$, $\quad$ = £216	M1 A1 M1, A1 (4) **[13]**
(a)	M1 for use of Po(2) may be implied A1 awrt 0.09	
(b)	B1 for Po(8) seen or used M1 for $1 - P(Y \le 10)$ oe A1 awrt 0.184	
(c)	1st M1 for forming a suitable Poisson distribution of the form $e^{-\lambda} = 0.8$ 1st A1 for use of lambda as $\dfrac{2x}{15}$ (this may appear after taking logs) 2nd M1 for attempt to consider a range of values that will prove 1.7 is correct **OR** for use of logs to show lambda = … 2nd A1 correct solution only. Either get 1.7 from using logs or stating values either side	
S.C	for $e^{-\frac{2}{15}\times 1.7} = 0.797... \approx 0.80 \quad \therefore x = 1.7$ to 2 sf allow 2nd M1A0	
(d)	1st M1 for one of the following 1200 p or 1200 (1 – p) where p = 0.8 or 2/15. 1st A1 for both expected values being correct or two correct expressions. 2nd M1 for an attempt to find expected profit, must consider with and without faults 2nd A1 correct answer only.	