Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Advanced Level

Thursday 26 May 2011 – Morning

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	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 7 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.

1. A factory produces components. Each component has a unique identity number and it is assumed that 2% of the components are faulty. On a particular day, a quality control manager wishes to take a random sample of 50 components.

(*a*) Identify a sampling frame.

(1)

The statistic F represents the number of faulty components in the random sample of size 50.

(*b*) Specify the sampling distribution of F.

(2)

---

2. A traffic officer monitors the rate at which vehicles pass a fixed point on a motorway. When the rate exceeds 36 vehicles per minute he must switch on some speed restrictions to improve traffic flow.

(*a*) Suggest a suitable model to describe the number of vehicles passing the fixed point in a 15 s interval.

(1)

The traffic officer records 12 vehicles passing the fixed point in a 15 s interval.

(*b*) Stating your hypotheses clearly, and using a 5% level of significance, test whether or not the traffic officer has sufficient evidence to switch on the speed restrictions.

(6)

(*c*) Using a 5% level of significance, determine the smallest number of vehicles the traffic officer must observe in a 10 s interval in order to have sufficient evidence to switch on the speed restrictions.

(3)

3.

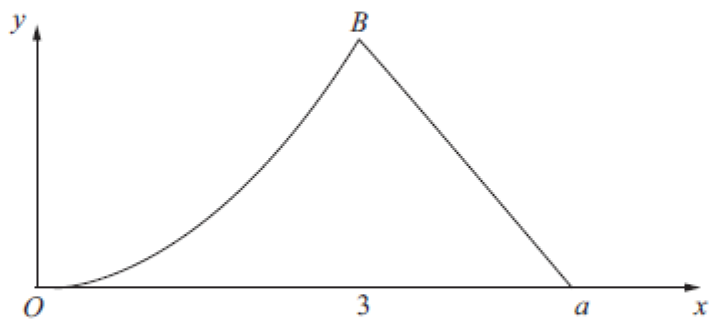

Figure 1

Figure 1 shows a sketch of the probability density function f(x) of the random variable X.

For $0 \leq x \leq 3$, f(x) is represented by a curve OB with equation f(x) = kx^2, where k is a constant.

For $3 \leq x \leq a$, where a is a constant, f(x) is represented by a straight line passing through B and the point (a, 0).

For all other values of x, f(x) = 0.

Given that the mode of X = the median of X, find

(*a*) the mode,

(1)

(*b*) the value of k,

(4)

(*c*) the value of a.

(3)

Without calculating E(X) and with reference to the skewness of the distribution

(*d*) state, giving your reason, whether E(X) < 3, E(X) = 3 or E(X) > 3.

(2)

4. In a game, players select sticks at random from a box containing a large number of sticks of different lengths. The length, in cm, of a randomly chosen stick has a continuous uniform distribution over the interval [7, 10].

A stick is selected at random from the box.

(*a*) Find the probability that the stick is shorter than 9.5 cm.

(2)

To win a bag of sweets, a player must select 3 sticks and wins if the length of the longest stick is more than 9.5 cm.

(*b*) Find the probability of winning a bag of sweets.

(2)

To win a soft toy, a player must select 6 sticks and wins the toy if more than four of the sticks are shorter than 7.6 cm.

(*c*) Find the probability of winning a soft toy.

(4)

---

5. Defects occur at random in planks of wood with a constant rate of 0.5 per 10 cm length. Jim buys a plank of length 100 cm.

(*a*) Find the probability that Jim's plank contains at most 3 defects.

(2)

Shivani buys 6 planks each of length 100 cm.

(*b*) Find the probability that fewer than 2 of Shivani's planks contain at most 3 defects.

(5)

(*c*) Using a suitable approximation, estimate the probability that the total number of defects on Shivani's 6 planks is less than 18.

(6)

---

6. A shopkeeper knows, from past records, that 15% of customers buy an item from the display next to the till. After a refurbishment of the shop, he takes a random sample of 30 customers and finds that only 1 customer has bought an item from the display next to the till.

(*a*) Stating your hypotheses clearly, and using a 5% level of significance, test whether or not there has been a change in the proportion of customers buying an item from the display next to the till.

(6)

During the refurbishment a new sandwich display was installed. Before the refurbishment 20% of customers bought sandwiches. The shopkeeper claims that the proportion of customers buying sandwiches has now increased. He selects a random sample of 120 customers and finds that 31 of them have bought sandwiches.

(*b*) Using a suitable approximation and stating your hypotheses clearly, test the shopkeeper's claim. Use a 10% level of significance.

(8)

---

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

$$f(x) = \begin{cases} \frac{3}{32}(x-1)(5-x) & 1 \leq x \leq 5, \\ 0 & \text{otherwise.} \end{cases}$$

(*a*) Sketch f(x) showing clearly the points where it meets the x-axis. **(2)**

(*b*) Write down the value of the mean, μ, of X. **(1)**

(*c*) Show that $E(X^2) = 9.8$. **(4)**

(*d*) Find the standard deviation, σ, of X. **(2)**

The cumulative distribution function of X is given by

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{32}(a - 15x + 9x^2 - x^3) & 1 \leq x \leq 5 \\ 1 & x > 5 \end{cases}$$

where a is a constant.

(*e*) Find the value of a. **(2)**

(*f*) Show that the lower quartile of X, q_1, lies between 2.29 and 2.31. **(3)**

(*g*) Hence find the upper quartile of X, giving your answer to 1 decimal place. **(1)**

(*h*) Find, to 2 decimal places, the value of k so that

$$P(\mu - k\sigma < X < \mu + k\sigma) = 0.5.$$

(2)

TOTAL FOR PAPER: 75 MARKS

END

June 2011
6684 Statistics S2
Mark Scheme

Question Number	Scheme	Marks
1. (a)	The list of ID numbers	B1 (1)
(b)	$F \sim B(50, 0.02)$	B1 B1 (2) **3**
Notes: (a)	B1 for idea of list/register/database and identity numbers	
	NB B0 if referring to the sample or 50 or only part of the population.	
(b)	**These must be in part (b) to gain the marks**	
	1st B1 for Binomial distribution	
	2nd B1 for $n = 50$ and $p = 0.02$ or (50,0.02)	
	NB (0.02, 50) is B0	
	Po(1) alone is B0B0	
	For a probability table	
	1st B1 Use of B(50,0.02) NB $P(X = 0) = 0.3642$	
	2nd B1 Table must have all 50 values and their probabilities.	

Question Number	Scheme	Marks
2. **(a)**	Poisson	B1 (1)
(b)	$H_0 : \mu = 9$ (or $\lambda = 36$) $H_1 : \mu > 9$ (or $\lambda > 36$)	B1 B1
	X ~Po(9) and P($X \geq 12$) = 1 - P($X \leq 11$) or P($X \leq 14$) = 0.9585; P($X \geq 15$) = 0.0415	M1
	= 1-0.8030 = 0.197 CR X $\geq$15	A1
	(0.197 > 0.05) so not significant/ accept H_0/ Not in CR	M1d
	he does not have evidence to switch on the speed restrictions (o.e)	A1ft (6)
(c)	Let Y = the number of vehicles in 10 s then Y ~Po(6)	B1
	Tables: P($Y \leq 10$) = 0.9574 so P($Y \geq 11$) = 0.0426	M1
	so needs 11 vehicles	A1 (3) **10**

Notes:

(a) B1 for Poisson or Po. Ignore their value for the mean.

(b) 1st B1 for $H_0 : \mu / \lambda = 9$ or $\mu / \lambda = 36$

2nd B1 for $H_1 : \mu / \lambda > 9$ or $\mu / \lambda > 36$

One tail

1st M1 for writing or using 1 - P($X \leq 11$) or writing P($X \leq 14$) = 0.9585 or P($X \geq 15$) = 0.0415. May be implied by correct CR.or probability = 0.197

A1 for 0.197 or a correct CR. Allow X >14. NB P($X \leq 11$) = 0.8030 on its own scores M1A1

2nd M1 dependent on the 1st M1 being awarded. For a correct statement based on the table below. Do not allow non-contextual conflicting statements eg "significant" and "accept H_0". **Ignore comparisons**.

2nd A1 for a correct contextualised statement. NB A correct contextual statement on its own scores M1A1.

	$0.05 < p < 0.95$	$p < 0.05$ or $p > 0.95$
2nd M1	not significant/ accept H_0/ Not in CR	significant/ reject H_0/ In CR
2nd A1	Insufficient evidence to switch on the speed restrictions	Sufficient evidence to switch on the speed restrictions

Two tail

1st M1 for writing or using 1 - P($X \leq 11$) or writing P($X \leq 15$) = 0.9780 or P($X \geq 16$) = 0.022. May be implied by correct CR. or probability = 0.197

A1 for 0.197 or CR $X \geq 16$. Allow X >15. NB P($X \leq 11$) = 0.8030 on its own scores M1A1

2nd M1 dependent on the 1st M1 being awarded . For a correct statement based on the table below. Do not allow non-contextual conflicting statements eg"significant" and "accept H_0" . **Ignore**

Question Number	Scheme	Marks
	comparisons. 2nd A1 for a correct contextualised statement. NB A correct contextual statement on its own scores M1A1. <table><tr><td></td><td>$0.025 < p < 0.975$</td><td>$p < 0.025$ or $p > 0.975$</td></tr><tr><td>2nd M1</td><td>not significant/ accept H_0/ Not in CR</td><td>significant/ reject H_0/ In CR</td></tr><tr><td>2nd A1</td><td>Insufficient evidence to switch on the speed restrictions</td><td>Sufficient evidence to switch on the speed restrictions</td></tr></table>	
(c)	B1 for identifying Po(6) - may be implied by use of correct tables M1 any one of the probs 0.9574 or 0.0426 or 0.9799 or 0.0201 may be implied by correct answer of 11 A1 cao do not accept $X \geq 11$ NB answer of 11 with no working gains all three marks.	
3. **(a)**	Mode = 3 from graph	B1 (1)
(b)	$\int_0^3 kx^2 \, dx = 0.5 \quad \Rightarrow \left[\frac{kx^3}{3}\right]_0^3 = 0.5$ So $\frac{27k}{3} - 0 = 0.5 \quad \Rightarrow k = \frac{1}{18}$ (using median = 3)	M1 A1 M1d A1 (4)
(c)	Height of triangle = $\frac{1}{18} \times 3^2 = \frac{1}{2}$ Area of triangle = $\frac{1}{2} \times (a-3) \times \frac{1}{2} = \frac{1}{2}$ so $a = 5$ cao	B1ft M1 A1 (3)
(d)	From graph distribution is negative skew (left tail is longer) $\mu <$ median for negative skew so E(X) < 3 [N.B. E(X) = $2\frac{23}{24}$]	B1 B1d (2) **10**
Notes: (b) (c)	1st M1 for attempt to integrate f(x) (need x^3). Integration must be in part (b) 1st A1 for correct integration. Ignore limits for these two marks. 2nd M1 Dependent on the previous M mark being awarded. For use of correct limits and set equal to 0.5 - leading to a linear equation for k. No need to see 0 substituted. 2nd A1 for $k = \frac{1}{18}$ or exact equivalent NB $k = \frac{1}{18}$ with no working gains M0A0M0A0 $k = \frac{1/2}{9} = \frac{1}{18}$ without sight of integration is M0A0M0A0 B1 for correct height of triangle using their k. ie $9k$. May be seen in working for area of triangle. Or correct gradient of line ie $\frac{9k}{(3-a)}$ o.e.	

Question Number	Scheme	Marks
	M1 for a correct linear equation for a, in the form $\pm\frac{1}{2}\times(a-3)\times 9k=\frac{1}{2}$ (Must see the halves) NB if they have stated their height and then used their height rather than 9k allow M1 A1 cao NB stating a = 5 and then verifying area of the triangle = 0.5 is acceptable. NB a = 5 on its own is B0M0A0 SC Integration of both parts = 1 or Integration of line = 0.5 leading to $a^2-8a+15=0$ gets B1 M1 and if they identify a = 5 A1	
(d)	1st B1 for identifying negative skew 2nd B1 dependent on previous B mark being awarded. For correct deduction E(X) <3	
4 (a)	$\frac{9.5-7}{10-7}$ $=\frac{5}{6}$ awrt 0.833	M1 A1 (2)
(b)	P(Longest > 9.5) = 1 - P(all < 9.5) = $1-\left(\frac{5}{6}\right)^3$ $=\frac{91}{216}$ or 0.421	M1 A1 (2)
(c)	P(a stick < 7.6) = $\frac{0.6}{3}=0.2$ Let Y = number of sticks (out of 6) <7.6 then Y~B(6, 0.2) P(Y > 4) = 1 - P($Y \le 4$) = 1 - 0.9984 = 0.0016 or $\frac{1}{625}$	B1 M1 M1 A1 (4) **8**
Notes: (a) (b) (c)	M1 for an expression for the probability e.g. $\int_7^{9.5}\frac{1}{3}\,dx$ M1 for $1-(a)^3$ **or** $(1-a)^3+3(1-a)^2a+3(1-a)a^2$ A1 awrt 0.421 B1 0.2 may be implied by at least one correct probability 1st M1 for writing or using B(6, p) may be implied by $np^x(1-p)^{6-x}$ using their p and $n \ge 1$ 2nd M1 for writing or using 1 – P($Y \le 4$) or $np^5(1-p)+p^6$ (n is an integer > 1) A1 cao NB 0.0016 with no working gets B0M0M0A0	
5. **(a)**	X ~Po(5); P($X \le 3$) = 0.2650	M1 A1 (2)

Question Number	Scheme	Marks
(b)	Let Y = the no.of planks with at most 3 defects, Y ~Binomial Y~B(6, 0.265) $P(Y < 2) = P(Y \leq 1)$ $= \left[0.735^6 + 6 \times 0.265 \times 0.735^5\right]$ = 0.4987…. **awrt 0.499 or 0.498**	M1 A1ft M1 A1 A1 (5)
(c)	Let T = total number of defects on 6 planks, T ~Po(30) so $T \approx S$ ~Normal S~N(30, 30) $P(T < 18) = P(S < 17.5)$ $= P\left(z < \frac{17.5 - 30}{\sqrt{30}}\right)$ $= P(Z < -2.28…)$ = 0.01123... **awrt 0.0112 or 0.0113**	M1 A1 M1 M1 A1 A1 (6) **13**
Notes: (a) (b) (c)	 M1 for identifying Po(5) - it should be clearly seen somewhere or implied A1 for correct probability. Allow 0.265 1st M1 for writing or using the binomial - may be implied by use of $nq^x(1-q)^{6-x}$ with $n \geq 1$ 1st A1ft for n = 6 and p = their (a) may be implied by $6p(1-p)^5$ or $(1-p)^6$ **NB** if they write B(6,(a)) they get M1 A1 2nd M1 for writing $P(Y \leq 1)$ **or** $P(Y = 0) + P(Y = 1)$ **or** $(1-q)^6 + nq(1-q)^5$ with $n \geq 1$ 2nd A1 $(1-p)^6 + 6p(1-p)^5$ where p = their (a) 3rd A1 for awrt 0.499 SC use of a probability in the tables – lose last two marks – could get M1A1M1 M0 A0 1st M1 for a normal approx 1st A1 for correct mean and sd 2nd M1 for use of continuity correction, either 17.5 or 18.5 or 42.5 or 41.5 seen 3rd M1 Standardising with their mean and their sd and 17.5 or 18 or 18.5 or 41.5 or 42 or 42.5 NB if they have not written down a mean and sd then they need to be correct in the standardisation to gain this mark. 2nd A1 for $z = \pm$ 2.28 or better. May be awarded for $\pm\frac{17.5-30}{\sqrt{30}}$ [NB no continuity correction z = 2.19] 3rd A1 for awrt 0.0112 or 0.0113 [NB no approximation gives 0.00727…] **SC using P(X<18.5) – P(X<17.5) can get M1 A1 M1 M0A0A0**	

Question Number	Scheme	Marks
6. **(a)**	$H_0 : p = 0.15 \quad H_1 : p \neq 0.15$	B1 B1
	$X \sim B(30, 0.15)$	M1
	$P(X \leq 1) = 0.0480$ or CR: $X = 0$	A1
	$(0.0480 > 0.025)$	
	not a significant result or do not reject H_0 or not in CR	M1
	there is no evidence of a change in the proportion of customers buying an item from the display.	A1ft
		(6)
(b)	$H_0 : p = 0.2 \quad H_1 : p > 0.2$	B1
	Let S = the number who buy sandwiches, $S \sim B(120, 0.2)$,	
	$S \approx W \sim N\left(24, \sqrt{19.2}^2\right)$	M1 A1
	$P(S \geq 31) = P(W \geq 30.5)$	M1
	$= P\left(Z > \frac{30.5 - 24}{\sqrt{19.2}}\right)$ or $\frac{x - 0.5 - 24}{\sqrt{19.2}} = 1.2816$	M1
	$[= P(Z > 1.48..)]$	
	$= 1 - 0.9306$	M1
	$= 0.0694$ $\quad x = 30.1$	A1
	< 0.10 so a significant result, there is evidence that more customers are purchasing sandwiches or the shopkeepers claim is correct.	B1ft
		(8)
Notes:		**14**

Notes:

(a)

1st B1 for H_0 must use p 2nd B1 for H_1 must use p

1st M1 for writing or using B(30,0.15) – may be implied by correct CR

1st A1 0.0480 or $X = 0$. Allow $X \leq 0$. Ignore upper CR. NB Allow CR $X \leq 1$ if using one tail test.

2nd M1 A correct statement (see table below) Do not allow non-contextual conflicting statements eg"significant" and "accept H_0". **Ignore comparisons**

2nd A1 for a correct statement in context. For context we need idea of change/decrease in number of customers buying from display – may use different words. NB A correct contextual statement on its own scores M1A1

	Two tail $0.025 < p < 0.975$ **or** **One tail** $0.05 < p < 0.95$	**Two tail** $p < 0.025$ or $p > 0.975$ **or** **One tail** $p < 0.05$ or $p > 0.95$
2nd M1	not significant/ accept H_0/ Not in CR or contextual	significant/ reject H_0/ In CR or contextual
2nd A1	There is no evidence of a change/decrease in the proportion of customers buying an item from the display	There is evidence of a change/decrease in the proportion of customers buying an item from the display.

(b)

1st B1 both hypotheses correct – must use p.

1st M1 for a normal approx

1st A1 for correct mean and sd

2nd M1 for use of continuity correction, either 30.5 or 31.5 or $(x \pm 0.5)$ seen

3rd M1 standardising with their mean and their sd and 30.5, 31 or 31.5 or x or $(x \pm 0.5)$)

4th M1 for 1 - tables value or 1.2816

2nd A1 for awrt 0.069 or $x = 30.1$

2nd B1ft For a correct conclusion in context using their probability and 0.1 For context we need idea of more customers buying sandwiches – may use different words

Question Number	Scheme	Marks
		One tail $0.1 < p < 0.9$ **or Two tail** $0.05 < p < 0.95$	**One tail** $p < 0.1$ or $p > 0.9$ **or Two tail** $p < 0.05$ or $p > 0.95$		
	2nd M1	not significant/ accept H_0/ Not in CR or contextual	significant/ reject H_0/ In CR or contextual	
	2nd A1	There is no evidence of an increase in the proportion of customers buying sandwiches	There is evidence of a change/increase in the proportion of customers buying sandwiches.	
	SC using $P(X<31.5) - P(X<30.5)$ can get B1M1 A1 M1 M1M0A0B0	
7 (a)	$\cap$ shape which does not go below the x-axis [condone missing patios] Graph must end at the points (1,0) and (5,0) and the points labelled at 1 and 5	B1 B1 (2)
(b)	E(X) = 3 (by symmetry)	B1 (1)
(c)	$\left[E(X^2)\right] = \int x^2 f(x)dx = \frac{3}{32}\int(6x^3 - x^4 - 5x^2)dx$ $= \frac{3}{32}\left[\frac{6x^4}{4} - \frac{x^5}{5} - \frac{5x^3}{3}\right]_1^5$ $= \frac{3}{32}\left(\left[\frac{6\times625}{4} - 625 - \frac{625}{3}\right] - \left[\frac{6}{4} - \frac{1}{5} - \frac{5}{3}\right]\right) = 9.8$ **(*)**	M1 A1 M1 A1 cso (4)
(d)	s.d. = $\sqrt{9.8 - E(X)^2}$, = 0.8944… **awrt 0.894**	M1 A1 (2)
(e)	F(1) = 0 $\Rightarrow \frac{1}{32}(a - 15 + 9 - 1) = 0$, leading to $\underline{a = 7}$	M1 A1 (2)
(f)	F(2.29) = 0.2449…, F(2.31) = 0.2515… Since F(q_1) = 0.25 and these values are either side of 0.25 then 2.29< q_1 < 2.31	M1 A1 A1 (3)
(g)	Since the distribution is symmetric q_3 =5-1.3 = $\underline{3.7}$ cao	B1 (1)
(h)	We know P(q_1=2.3 < X < 3.7=q_3) = 0.5 so $k\sigma = 0.7$ so $k = \frac{0.7}{0.894...} = 0.7826..$ = **awrt 0.78**	M1 A1 (2) **17**

Question Number	Scheme	Marks
Notes: (c)	This part is a "show that" therefore we need to see all the steps in the working 1st M1 for showing intention of doing $\int x^2 \mathrm{f}(x)$ and attempt to multiply out bracket 1st A1 for correct integration, cao, ignore limits for this mark. 2nd M1 for use of correct limits. Need to see evidence of subst both 5 and 1. 2nd A1 for cso leading to 9.8. Do not ignore subsequent working for this final A mark.	
(d)	M1 for a correct expression for standard deviation, must include $\sqrt{\ldots}$ A1 allow awrt 0.894, $\sqrt{0.8}$, $\frac{2\sqrt{5}}{5}$ oe	
(e)	M1 for a correct method to find a. e.g F(5) = 1 or $\int_1^5 f(x) = 1$	
(f)	M1 for an attempt at F(2.29) or F(2.31) or put F(x) = 0.25 (ft their value of a) 1st A1 for both values seen. awrt 0.245 and 0.252 find 3 solutions awrt 6.76/6.75, 2.305, -0.064 2nd A1 for comparison with 0.25 and stating Q_1 state only 2.30 in range and stating Q_1 lies between 2.29 and 2.31 lies between 2.29 and 2.31	
(h)	M1 For $k\sigma$ = awrt 0.7 A1 Allow awrt 0.78 NB a correct awrt 0.78 gains M1 A1	