Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Bronze Level B4

Time: 1 hour 30 minutes

Materials required for examination papers	**Items included with question**
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

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, initials and signature.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
There are 6 questions in this question 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.

Suggested grade boundaries for this paper:

A*	A	B	C	D	E
70	65	58	51	46	40

1. A teacher is monitoring the progress of students using a computer based revision course. The improvement in performance, y marks, is recorded for each student along with the time, x hours, that the student spent using the revision course. The results for a random sample of 10 students are recorded below.

x hours	1.0	3.5	4.0	1.5	1.3	0.5	1.8	2.5	2.3	3.0
y marks	5	30	27	10	–3	–5	7	15	–10	20

[You may use $\sum x = 21.4$, $\sum y = 96$, $\sum x^2 = 57.22$, $\sum xy = 313.7$]

(*a*) Calculate S_{xx} and S_{xy}.

(3)

(*b*) Find the equation of the least squares regression line of y on x in the form $y = a + bx$.

(4)

(*c*) Give an interpretation of the gradient of your regression line.

(1)

Rosemary spends 3.3 hours using the revision course.

(*d*) Predict her improvement in marks.

(2)

Lee spends 8 hours using the revision course claiming that this should give him an improvement in performance of over 60 marks.

(*e*) Comment on Lee's claim.

(1)

January 2009

---

2. The random variable X has probability distribution given in the table below.

x	–1	0	1	2	3
P($X = x$)	p	q	0.2	0.15	0.15

Given that E(X) = 0.55, find

(*a*) the value of p and the value of q,

(5)

(*b*) Var (X),

(4)

(*c*) E($2X – 4$).

(2)

May 2008

---

3. A student is investigating the relationship between the price (y pence) of 100g of chocolate and the percentage (x%) of the cocoa solids in the chocolate.
The following data is obtained

Chocolate brand	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
x (% cocoa)	10	20	30	35	40	50	60	70
y (pence)	35	55	40	100	60	90	110	130

(You may use: $\sum x = 315,\ \sum x^2 = 15\,225,\ \sum y = 620,\ \sum y^2 = 56\,550,\ \sum xy = 28\,750$)

(*a*) Draw a scatter diagram to represent these data.

(2)

(*b*) Show that $S_{xy} = 4337.5$ and find S_{xx}.

(3)

The student believes that a linear relationship of the form $y = a + bx$ could be used to describe these data.

(*c*) Use linear regression to find the value of a and the value of b, giving your answers to 1 decimal place.

(4)

(*d*) Draw the regression line on your diagram.

(2)

The student believes that one brand of chocolate is overpriced.

(*e*) Use the scatter diagram to

(i) state which brand is overpriced,

(ii) suggest a fair price for this brand.

Give reasons for both your answers.

(4)

June 2007

4. The discrete random variable X has probability function

$$P(X = x) = \begin{cases} a(3-x) & x = 0, 1, 2 \\ b & x = 3 \end{cases}$$

(a) Find $P(X = 2)$ and copy and complete the table below.

x	0	1	2	3
$P(X=x)$	$3a$	$2a$		b

(1)

Given that $E(X) = 1.6$,

(b) find the value of a and the value of b.

(5)

Find

(c) $P(0.5 < X < 3)$,

(2)

(d) $E(3X - 2)$.

(2)

(e) Show that the $Var(X) = 1.64$

(3)

(f) Calculate $Var(3X - 2)$.

(2)

May 2009

---

5. The random variable X has a normal distribution with mean 30 and standard deviation 5.

(a) Find $P(X < 39)$.

(2)

(b) Find the value of d such that $P(X < d) = 0.1151$.

(4)

(c) Find the value of e such that $P(X > e) = 0.1151$.

(2)

(d) Find $P(d < X < e)$.

(2)

January 2009

---

6. The random variable X has probability distribution

x	1	3	5	7	9
$P(X = x)$	0.2	p	0.2	q	0.15

(a) Given that $E(X) = 4.5$, write down two equations involving p and q.

(3)

Find

(b) the value of p and the value of q,

(3)

(c) $P(4 < X \leq 7)$.

(2)

Given that $E(X^2) = 27.4$, find

(d) Var (X),

(2)

(e) $E(19 - 4X)$,

(1)

(f) Var $(19 - 4X)$.

(2)

June 2007

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1. (a)	$S_{xx} = 57.22 - \frac{(21.4)^2}{10} = 11.424$	M1 A1
	$S_{xy} = 313.7 - \frac{21.4 \times 96}{10} = 108.26$	A1
		(3)
(b)	$b = \frac{S_{xy}}{S_{xx}} = 9.4765...$	M1 A1
	$a = \bar{y} - b\bar{x} = 9.6 - 2.14b = (-10.679...)$	M1
	$y = -10.7 + 9.48x$	A1
		(4)
(c)	Every (extra) <u>hour</u> spent using the programme produces about <u>9.5 marks improvement</u>	B1ft
		(1)
(d)	$y = -10.7 + 9.48 \times 3.3, = 20.6$ awrt 21	M1,A1
		(2)
(e)	Model may not be valid since [8h is] outside the range [0.5 - 4].	B1
		(1) **[11]**
2. (a)	$-1 \times p + 1 \times 0.2 + 2 \times 0.15 + 3 \times 0.15 \quad = 0.55$	M1 M1
	$p = 0.4$	A1
	$p + q + 0.2 + 0.15 + 0.15 \quad = 1$	M1
	$q = 0.1$	A1
		(5)
(b)	$\text{Var}(X) = (-1)^2 \times p + 1^2 \times 0.2 + 2^2 \times 0.15 + 3^2 \times 0.15, -0.55^2$	M1A1, M1
	$= 2.55 - 0.3025 = 2.2475$ awrt 2.25	A1
		(4)
(c)	$E(2X-4) = 2E(X)-4$	M1
	$= -2.9$	A1
		(2) **[11]**

Question Number	Scheme	Marks
3. (a)	Use overlay	B2 (2)
(b)	$S_{xy} = 28750 - \frac{315 \times 620}{8} = 4337.5$ **answer given**	M1
	$S_{xx} = 15225 - \frac{315^2}{8} = 2821.875$	M1A1 (3)
(c)	$b = \frac{4377.5}{S_{xx}}, = 1.537... = 1.5$	M1,A1
	$a = \overline{y} - b\overline{x} = \frac{620}{8} - b\frac{315}{8} = 16.97... = 17.0$	M1,A1 (4)
(d)	Use overlay	B1∫ B1 (2)
(e)	Brand D,	B1
	since a long way above / from the line dependent upon 'Brand D' above	B1
	Using line: $y = 17 + 35 \times 1.5 = 69.5$	M1A1 (4) **[15]**

Question Number	Scheme		Marks
4. (a)	<table><tr><td>0</td><td>1</td><td>2</td><td>3</td></tr><tr><td>3a</td><td>2a</td><td>a</td><td>b</td></tr></table>		B1 (1)
(b)	$3a + 2a + a + b = 1$	or equivalent	M1
	$2a + 2a + 3b = 1.6$	or equivalent	M1
	$14a = 1.4$		M1
	$a = 0.1$	cao	B1
	$b = 0.4$	cao	B1 (5)
(c)	$P(0.5 < x < 3) = P(1) + P(2)$		M1
	$= 0.2 + 0.1$		
	$= 0.3$		A1 ft (2)
(d)	$E(3X - 2) = 3E(X) - 2$		M1
	$= 3 \times 1.6 - 2$		
	$= 2.8$	cao	A1 (2)
(e)	$E(X^2) = 1\times0.2 + 4 \times0.1 + 9 \times0.4$ (= 4.2)		M1
	Var (X) = "4.2" – 1.6^2		M1
	= 1.64 **given answer**	cso	A1 (3)
(f)	$\text{Var}(3X - 2) = 9\ \text{Var}(X)$		M1
	= 14.76	awrt 14.8	A1 (2) **[15]**

Question Number	Scheme	Marks
5. (a)	$P(X < 39) = P\left(Z < \frac{39-30}{5}\right)$	M1
	$= P(Z < 1.8) = \underline{0.9641}$ (allow awrt 0.964)	A1
		(2)
(b)	$P(X < d) = P\left(Z < \frac{d-30}{5}\right) = 0.1151$	M1
	$1 - 0.1151 = 0.8849$	
	$\Rightarrow \quad z = -1.2$ (allow $\pm$ 1.2)	B1
	$\therefore \frac{d-30}{5} = -1.2$ $\underline{d = 24}$	M1A1
		(4)
(c)	$P(X > e) = 0.1151$ so $e = \mu + (\mu$ - their $d)$ or	
	$\frac{e-30}{5} = 1.2$ or $-$ their z	M1
	$\underline{e = 36}$	A1
		(2)
(d)	$P(d < X < e) = 1 - 2 \times 0.1151$	M1
	$= 0.7698$ awrt $\underline{0.770}$	A1
		(2)
		[10]
6. (a)	$p + q = 0.45$	B1
	$\sum xP(X = x) = 4.5$	M1
	$3p + 7q = 1.95$	A1
		(3)
(b)	Attempt to solve equations in (a)	M1
	$q = 0.15$	A1
	$p = 0.30$	A1
		(3)
(c)	$P(4 < X < 7) = P(5) + P(7)$	M1
	$= 0.2 + q = 0.35$	A1$\int$
		(2)
(d)	$\text{Var}(X) = E(X^2) - [E(X)]^2 = 27.4 - 4.5^2$	M1
	$= 7.15$	A1
		(2)
(e)	$E(19 - 4X) = 19 - 4 \times 4.5 = 1$	B1
		(1)
(f)	$\text{Var}(19 - 4X) = 16\text{Var}(X)$	M1
	$= 16 \times 7.15 = 114.4$	A1
		(2)
		[13]

Examiner reports

Question 1

This proved to be a straightforward starter for most candidates who were able to tackle part (a) confidently, usually scoring full marks. Part (b) was answered well too; the correct formulae were selected and answers were usually given to 3 sf or better. Some candidates lost the final mark here for failing to give the full equation. Part (c) though was not answered well. There were plenty of comments about the gradient being positive or there being positive correlation or even skewness. Few realised that the instruction to "interpret" wanted an answer in context and comments conveying the idea that every extra hour spent on the programme yields an extra 9.5 marks were rare. Part (d) was straightforward again but some did not use their regression equation to find the estimate but rather tried to interpolate between the values of 3 and 3.5 given in the table. Part (e) had a mixed response. Many good candidates rejected Lee's comment on the basis that 8 hours was outside the range of the data and they secured the mark. Other, less successful, candidates simply calculated the value and then agreed with Lee or they rejected his claim on some other basis such as the difficulty of revising for 8 hours or 60 marks might take him above the total score on the paper.

Question 2

This proved to be a good question allowing the better students to show their understanding of the topic. It was generally done well with the majority of students aware of what they were trying to achieve. Of those who did less well, many failed to realise the significance of the 0.55 and others only came up with one equation in part (a) and thought that q was therefore equal to 0. A very large number of candidates incorrectly squared -1 in part (b) affecting their calculation of variance. Part (c) was generally well answered.

Question 3

This was generally well answered. The majority of the errors occurred in part (c) by rounding too early and getting 18.4 for a. The regression line was often inaccurately plotted. In part (e) many used chocolate content to justify answer and often did not use the regression line to get a suitable price. Some misunderstood the question and attempted to find the best value in the second part of part (e).

Question 4

This entire question was usually very well done if part (b) was correct. Some candidates did not identify that the sum of probabilities should equal one and had problems trying to find the values of a and b resorting to guessing. Even if candidates could form two correct equations, some lacked the ability to solve these relatively simple equations. A number of candidates who had no success with part (b) gave up at this point but others managed to get part (d) and part (f) correct using the values given in the question. In part (e) many knew that they had to take 1.6^2 from their figure, and not having got the figures for a and b correct in part (b) they adjusted their number to come to 4.2, so that $4.2 - 1.6^2$ came to 1.64. On occasion it was difficult to distinguish between $\sum$ and E in the candidate's handwriting.

Question 5

Most candidates tried this question and the standardisation in part (a) was usually correct but a small minority used 25 as the standard deviation. The majority found P($Z < 1.8$) correctly but some gave the answer as 1 - 0.9641 and lost the second mark.

A clear diagram should have helped candidates with the next two parts for many gave answers to d and e where $d > e$. In part (b) many started correctly by calculating 1 - 0.1151 and using the tables to find $z = \pm 1.2$. However only the more alert chose the minus sign and they usually went on to score full marks in both parts (b) and (c). There were good arguments using the symmetry of the normal distribution in both parts (c) and (d). Some candidates who made little progress with (b) or (c) were able to draw a simple diagram in (d) and obtain the correct answer from $1 - 2 \times 0.1151$.

Question 6

A sizeable minority of candidates did not attempt this question, but, when attempted, this question was well done with many candidates picking up most or all the marks. Rather surprisingly in part (a) the equation missing was $p + q = 0.45$ and a few candidates divided one side of their second equation by 5. In part (b) nearly all of those who had two correct equations for part (a) were able to solve them simultaneously. In part (c) a substantial number of candidates were unable to make a successful attempt at this part of the question with many omitting it entirely. There were a large number of accurate solutions to part (d) with most of those making an error failing to use the given values. A number of candidates reworked E(X^2) for part (e) even though it was given. There were some mistakes in part (f) but most candidates used 16 correctly, but some multiplied E(X) instead of Var(X).

Statistics for S1 Practice Paper Bronze Level B4

				Mean score for students achieving grade:						
Qu	Max Score	Modal score	Mean %	ALL	A*	A	B	C	D	E	U
1	11		78	8.56		9.56	8.77	8.11	7.62	6.87	4.57
2	11		78	8.62		10.63	10.15	9.47	8.15	6.36	3.19
3	15		77	11.58		13.60	12.66	11.98	11.17	10.13	7.18
4	15		73	10.97		14.42	13.43	11.93	9.76	7.12	3.29
5	10		64	6.40		8.89	6.87	4.88	3.75	3.11	1.07
6	13		67	8.69		12.21	10.96	9.42	7.27	5.21	1.96
	75		**73**	**54.82**		**69.31**	**62.84**	**55.79**	**47.72**	**38.80**	**21.26**