Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced Subsidiary

Monday 19 January 2009 – Afternoon

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 S1), the paper reference (6683), 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 6 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.

N32680A

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)

---

2. A group of office workers were questioned for a health magazine and $\frac{2}{5}$ were found to take regular exercise. When questioned about their eating habits $\frac{2}{3}$ said they always eat breakfast and, of those who always eat breakfast $\frac{9}{25}$ also took regular exercise.

Find the probability that a randomly selected member of the group

(a) always eats breakfast and takes regular exercise,

(2)

(b) does not always eat breakfast and does not take regular exercise.

(4)

(c) Determine, giving your reason, whether or not always eating breakfast and taking regular exercise are statistically independent.

(2)

---

3. When Rohit plays a game, the number of points he receives is given by the discrete random variable X with the following probability distribution.

x	0	1	2	3
$P(X = x)$	0.4	0.3	0.2	0.1

(*a*) Find E(X).

(2)

(*b*) Find F(1.5).

(2)

(*c*) Show that Var(X) = 1.

(4)

(*d*) Find Var(5 – 3X).

(2)

Rohit can win a prize if the total number of points he has scored after 5 games is at least 10. After 3 games he has a total of 6 points. You may assume that games are independent.

(*e*) Find the probability that Rohit wins the prize.

(6)

---

4. In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

The total length of calls, y minutes, for the 11 students were

17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110

(*a*) Find the median and quartiles for these data.

(3)

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(*b*) Show that 110 is the only outlier.

(2)

(*c*) Draw a box plot for these data indicating clearly the position of the outlier.

(3)

The value of 110 is omitted.

(*d*) Show that S_{yy} for the remaining 10 students is 2966.9

(3)

These 10 students were each asked how many text messages, x, they sent in the same week. The values of S_{xx} and S_{xy} for these 10 students are $S_{xx} = 3463.6$ and $S_{xy} = -18.3$.

(*e*) Calculate the product moment correlation coefficient between the number of text messages sent and the total length of calls for these 10 students.

(2)

A parent believes that a student who sends a large number of text messages will spend fewer minutes on calls.

(*f*) Comment on this belief in the light of your calculation in part (*e*).

(1)

---

5. In a shopping survey a random sample of 104 teenagers were asked how many hours, to the nearest hour, they spent shopping in the last month. The results are summarised in the table below.

Number of hours	Mid-point	Frequency
0 – 5	2.75	20
6 – 7	6.5	16
8 – 10	9	18
11 – 15	13	25
16 – 25	20.5	15
26 – 50	38	10

A histogram was drawn and the group (8 – 10) hours was represented by a rectangle that was 1.5 cm wide and 3 cm high.

(*a*) Calculate the width and height of the rectangle representing the group (16 – 25) hours.

(3)

(*b*) Use linear interpolation to estimate the median and interquartile range.

(5)

(*c*) Estimate the mean and standard deviation of the number of hours spent shopping.

(4)

(*d*) State, giving a reason, the skewness of these data.

(2)

(*e*) State, giving a reason, which average and measure of dispersion you would recommend to use to summarise these data.

(2)

---

6. 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)

TOTAL FOR PAPER: 75 MARKS

END

January 2009
6683 Statistics S1
Mark Scheme

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 M1
	$a = \bar{y} - b\bar{x} = 9.6 - 2.14b = (-10.679...)$	A1 (4)
(c)	$y = -10.7 + 9.48x$ Every (extra) hour spent using the programme produces about 9.5 marks improvement	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]**

(a) M1 for a correct expression
1st A1 for AWRT 11.4 for S_{xx}
2nd A1 for AWRT 108 for S_{xy}
Correct answers only: One value correct scores M1 and appropriate A1, both correct M1A1A1

(b)
1st M1 for using their values in correct formula
1st A1 for AWRT 9.5
2nd M1 for correct method for *a* (minus sign required)
2nd A1 for equation with *a* and *b* AWRT 3 sf (e.g. $y = -10.68 + 9.48x$ is fine)
Must have a full equation with *a* and *b* correct to awrt 3 sf

(c)
B1ft for comment conveying the idea of *b* marks per hour. Must mention value of *b* but can ft their value of *b*. No need to mention "extra" but must mention "marks" and "hour(s)" e.g. " …9.5 times per hour …" scores B0

(d)
M1 for sub $x = 3.3$ into their regression equation from the end of part (b)
A1 for awrt 21

(e) B1 for a statement that says or implies that it may not be valid because outside the range. They do not have to mention the values concerned here namely 8 h or 0.5 - 4

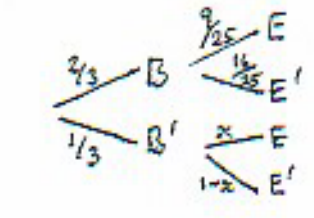
Question Number	Scheme	Marks
2 (a)	E = take regular exercise B = always eat breakfast $\mathrm{P}(E \cap B) = \mathrm{P}(E \mid B) \times \mathrm{P}(B)$ $= \frac{9}{25} \times \frac{2}{3} = 0.24$ or $\frac{6}{25}$ or $\frac{18}{75}$	M1 A1 (2)
(b)	$\mathrm{P}(E \cup B) = \frac{2}{3} + \frac{2}{5} - \frac{6}{25} = \frac{62}{75}$ | or $\mathrm{P}(E' \mid B') = \frac{13}{25}$ | or $\mathrm{P}(B' \cap E) = \frac{12}{75}$ | or $\mathrm{P}(B \cap E') = \frac{32}{75}$ $\mathrm{P}(E' \cap B') = 1 - \mathrm{P}(E \cup B) = \frac{13}{75}$ or $0.17\dot{3}$	M1 A1 M1 A1 (4)
(c)	$\mathrm{P}(E \mid B) = 0.36 \neq 0.40 = \mathrm{P}(E)$ or $\mathrm{P}(E \cap B) = \frac{6}{25} \neq \frac{2}{5} \times \frac{2}{3} = \mathrm{P}(E) \times \mathrm{P}(B)$ So E and B are <u>not</u> statistically independent	M1 A1 (2) **[8]**

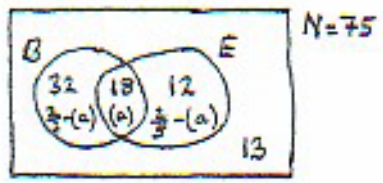
(a) M1 for $\frac{9}{25} \times \frac{2}{3}$ or $\mathrm{P}(E|B) \times \mathrm{P}(B)$ <u>and</u> at least one correct value seen. A1 for 0.24 or exact equiv.
NB $\frac{2}{5} \times \frac{2}{3}$ alone or $\frac{2}{5} \times \frac{9}{25}$ alone scores M0A0. Correct answer scores full marks.

(b) 1st M1 for use of the addition rule. Must have 3 terms and some values, can ft their (a)
<u>Or</u> a full method for $\mathrm{P}(E'|B')$ requires 1 - $\mathrm{P}(E|B')$ and equation for $\mathrm{P}(E|B')$: (a) $+\frac{x}{3} = \frac{2}{5}$
<u>Or</u> a full method for $\mathrm{P}(B' \cap E)$ <u>or</u> $\mathrm{P}(B \cap E')$ [or other valid method]
2nd M1 for a method leading to answer e.g. $1 - \mathrm{P}(E \cup B)$
<u>or</u> $\mathrm{P}(B') \times \mathrm{P}(E' \mid B')$ <u>or</u> $\mathrm{P}(B') - \mathrm{P}(B' \cap E)$ <u>or</u> $\mathrm{P}(E') - \mathrm{P}(B \cap E')$
<u>Venn Diagram</u> 1st M1 for diagram with attempt at $\frac{2}{5} - \mathrm{P}(B \cap E)$ or $\frac{2}{3} - \mathrm{P}(B \cap E)$. Can ft their (a)
1st A1 for a correct first probability as listed or 32, 18 and 12 on Venn Diagram
2nd M1 for attempting 75 - their (18 +32 + 12)

(c) M1 for identifying suitable values to test for independence e.g. $\mathrm{P}(E) = 0.40$ and $\mathrm{P}(E|B) = 0.36$
<u>Or</u> $\mathrm{P}(E) \times \mathrm{P}(B) = \ldots$ and $\mathrm{P}(E \cap B) =$ their (a) [but their (a) $\neq \frac{2}{5} \times \frac{2}{3}$]. Values seen somewhere
A1 for correct values and a correct comment

<u>Diagrams</u> You may see these or find these useful for identifying probabilities.





<u>Common Errors</u>
(a) $\frac{9}{25}$ is M0A0
(b) $\mathrm{P}(E \cup B) = \frac{53}{75}$ scores M1A0
$1 - \mathrm{P}(E \cup B) = \frac{22}{75}$ scores M1A0
(b) $\mathrm{P}(B') \times \mathrm{P}(E') = \frac{1}{3} \times \frac{3}{5}$
scores 0/4

Question Number	Scheme	Marks
3 (a)	$E(X) = 0\times0.4+1\times0.3+...+3\times0.1, \quad = \quad 1$	M1, A1 (2)
(b)	$F(1.5) = [P(X \le 1.5) =] \; P(X \le 1), \quad = 0.4 + 0.3 = 0.7$	M1, A1 (2)
(c)	$E(X^2) = 0^2\times0.4+1^2\times0.3+...+3^2\times0.1 \quad , = \; 2$ $Var(X) = \; 2-1^2 \; , \; = \; 1 \qquad (*)$	M1, A1 M1, A1cso (4)
(d)	$Var(5-3X) = (-3)^2 \, Var(X), \quad = 9$	M1, A1 (2)
(e)	<table><tr><th>Total</th><th>Cases</th><th>Probability</th></tr><tr><td rowspan="3">4</td><td>$(X=3)\cap(X=1)$</td><td>$0.1\times0.3=0.03$</td></tr><tr><td>$(X=1)\cap(X=3)$</td><td>$0.3\times0.1=0.03$</td></tr><tr><td>$(X=2)\cap(X=2)$</td><td>$0.2\times0.2=0.04$</td></tr><tr><td rowspan="2">5</td><td>$(X=3)\cap(X=2)$</td><td>$0.1\times0.2=0.02$</td></tr><tr><td>$(X=2)\cap(X=3)$</td><td>$0.2\times0.1=0.02$</td></tr><tr><td>6</td><td>$(X=3)\cap(X=3)$</td><td>$0.1\times0.1=0.01$</td></tr></table> Total probability = 0.03 + 0.03+0.04 +0.02 + 0.02 + 0.01 = 0.15	B1B1B1 M1 A1 A1 (6) **[16]**

(a) M1 for at least 3 terms seen. Correct answer only scores M1A1. Dividing by $k(\neq 1)$ is M0.

(b) M1 for F(1.5) = P($X \le 1$).[**Beware**: $2\times0.2+3\times0.1=0.7$ but scores M0A0]

(c) 1st M1 for at least 2 non-zero terms seen. $E(X^2) = 2$ alone is M0. Condone calling $E(X^2)$ =Var(X).
1st A1 is for an answer of 2 or a fully correct expression.

ALT
2nd M1 for $-\mu^2$, condone 2 – 1, unless clearly 2 - □. Allow $2-\mu^2$, with □ = 1 even if E(X) ≠ 1
2nd A1 for a fully correct solution with no incorrect working seen, **both** Ms required.

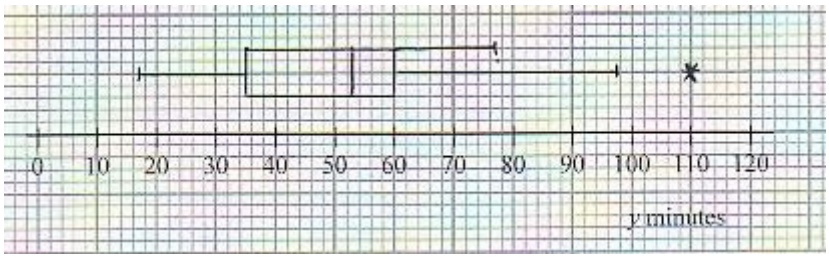
$\sum(x-\mu)^2 \times P(X=x)$

1st M1 for an attempt at a full list of $(x-\mu)^2$ values and probabilities. 1st A1 if all correct
2nd M1 for at least 2 non-zero terms of $(x-\mu)^2 \times P(X=x)$ seen. 2nd A1 for 0.4 + 0.2 + 0.4 = 1

(d) M1 for use of the correct formula. -3^2 Var(X) is M0 unless the final answer is >0.

(e) Can follow through their Var(X) for M1

ALT
1st B1 for all cases listed for a total of 4 or 5 or 6 . e.g. (2,2) counted twice for a total of 4 is B0
2nd B1 for all cases listed for 2 totals }
3rd B1 for a complete list of all 6 cases }These may be highlighted in a table
Using Cumulative probabilities
1st B1 for one or more cumulative probabilities used e.g.2 then 2 or more or 3 then 1 or more
2nd B1 for both cumulative probabilities used. 3rd B1 for a complete list 1, 3; 2, ≥2; 3, ≥1
M1 for one correct pair of correct probabilities multiplied
1st A1 for all 6 correct probabilities listed (0.03, 0.03, 0.04, 0.02, 0.02, 0.01) needn't be added.
2nd A1 for 0.15 or exact equivalent only as the final answer.

Question Number		Scheme	Marks
4	(a)	$Q_2 = 53, \quad Q_1 = 35, \quad Q_3 = 60$	B1, B1,B1 (3)
	(b)	$Q_3 - Q_1 = 25 \Rightarrow Q_1 - 1.5\times 25 = -2.5$ (no outlier) $Q_3 + 1.5\times 25 = 97.5$ (so 110 is an outlier)	M1 A1 (2)
	(c)		M1 A1ft A1ft (3)
	(d)	$\sum y = 461, \sum y^2 = 24\ 219 \therefore\ S_{yy} = 24219 - \frac{461^2}{10}$, $= 2966.9$ (*)	B1, B1, B1cso (3)
	(e)	$r = \frac{-18.3}{\sqrt{3463.6\times 2966.9}}$ or $\frac{-18.3}{3205.64...} = -0.0057$ AWRT - 0.006 or -6×10^{-3}	M1 A1 (2)
	(f)	*r* suggests correlation is close to zero so parent's claim is not justified	B1 (1) **[14]**

(a)	1st B1 for median 2nd B1 for lower quartile 3rd B1 for upper quartile
(b)	M1 for attempt to find one limit A1 for both limits found and correct. No explicit comment about outliers needed.
(c)	M1 for a box and two whiskers 1st A1ft for correct position of box, median and quartiles. Follow through their values. 2nd A1ft for 17 and 77 or "their" 97.5 and *. If 110 is not an outlier then score A0 here. Penalise no gap between end of whisker and outlier. Must label outlier, needn't be with *. **Accuracy** should be within the correct square so 97 or 98 will do for 97.5
(d)	1st B1 for $\sum y$ N.B. $\left(\sum y\right)^2 = 212521$ and can imply this mark 2nd B1 for $\sum y^2$ or at least three correct terms of $\sum\left(y-\bar{y}\right)^2$ seen. 3rd B1 for complete correct expression seen leading to 2966.9. So all 10 terms of $\sum\left(y-\bar{y}\right)^2$
(e)	M1 for attempt at correct expression for *r*. Can ft their S_{yy} for M1.
(f)	B1 for comment rejecting parent's claim on basis of weak or zero correlation Typical error is "negative correlation so comment is true" which scores B0 Weak negative or weak positive correlation is OK as the basis for their rejection.

Question Number	Scheme	Marks
5 (a)	8-10 hours: width = 10.5 - 7.5 = 3 represented by 1.5cm 16-25 hours: width = 25.5 - 15.5 = 10 so represented by 5 cm 8- 10 hours: height = fd = 18/3 = 6 represented by 3 cm 16-25 hours: height = fd = 15/10 = 1.5 represented by 0.75 cm	B1 M1 A1 (3)
(b)	$Q_2 = 7.5 + \frac{(52-36)}{18} \times 3 = 10.2$ $Q_1 = 5.5 + \frac{(26-20)}{16} \times 2 [= 6.25 \text{ or } 6.3]$ or $5.5 + \frac{(26.25-20)}{16} \times 2$ [=6.3] $Q_3 = 10.5 + \frac{(78-54)}{25} \times 5 [= 15.3]$ or $10.5 + \frac{(78.75-54)}{25} \times 5$ [=15.45 \15.5] IQR = (15.3 - 6.3) = 9	M1 A1 A1 A1 A1ft (5)
(c)	$\sum fx = 1333.5 \Rightarrow \bar{x} = \frac{1333.5}{104} =$ AWRT 12.8	M1 A1
(d)	$\sum fx^2 = 27254 \Rightarrow \sigma_x = \sqrt{\frac{27254}{104} - \bar{x}^2} = \sqrt{262.05 - \bar{x}^2}$ AWRT 9.88	M1 A1 (4)
(e)	$Q_3 - Q_2 [= 5.1] > Q_2 - Q_1 [= 3.9]$ or $Q_2 < \bar{x}$ So data is positively skew	B1ft dB1 (2)
	Use median and IQR, since data is skewed or not affected by extreme values or outliers	B1 B1 (2) **[16]**

(a) M1 For attempting both frequency densities $\frac{18}{3}(=6)$ and $\frac{15}{10}$, and $\frac{15}{10} \times \text{SF}$, where SF $\neq 1$

(b) NB Wrong class widths(2 and 9) gives $\frac{h}{1.66...} = \frac{3}{9} \rightarrow h = \frac{5}{9}$ or 0.55… and scores M1A0

(c) M1 for identifying correct interval and a correct fraction e.g. $\frac{\frac{1}{2}(104)-36}{18}$. Condone 52.5 or 53

1st A1 for 10.2 for median. Using (n + 1) allow awrt 10.3

NB: Must see some method

2nd A1 for a correct expression for either Q_1 or Q_3 (allow 26.25 and 78.75)

3rd A1 for correct expressions for both Q_1 and Q_3

4th A1ft for IQR, ft their quartiles. Using (n + 1) gives 6.28 and 15.45

(d) 1st M1 for attempting $\sum fx$ and $\bar{x}$

2nd M1 for attempting $\sum fx^2$ and $\sigma_x, \sqrt{}$ is needed for M1. Allow s = awrt 9.93

(e) 1st B1ft for suitable test, values need not be seen but statement must be compatible with values used. Follow through their values

2nd dB1 Dependent upon their test showing positive and for stating positive skew
If their test shows negative skew they can score 1st B1 but lose the second

1st B1 for choosing median and IQR. Must mention both. } Award independently

2nd B1 for suitable reason }

e.g. "use median because data is skewed" scores B0B1 since IQR is not mentioned

Question Number	Scheme	Marks
6 (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)	$P(X < d) = P\left(Z < \frac{d-30}{5}\right) = 0.1151$	
	$1 - 0.1151 = 0.8849$	M1
	$\Rightarrow \quad z = -1.2$ (allow $\pm$ 1.2)	B1
	$\therefore \frac{d-30}{5} = -1.2 \qquad \underline{d = 24}$	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]

Answer only scores all marks in each section BUT check (b) and (c) are in correct order

(a)

M1 for standardising with σ, $z = \pm\frac{39-30}{5}$ is OK

A1 for 0.9641 or awrt 0.964 but if they go on to calculate 1 – 0.9641 they get M1A0

(b)

1st M1 for attempting 1- 0.1151. Must be seen in (b) in connection with finding d

B1 for $z = \pm$ 1.2. They must state $z = \pm$ 1.2 or imply it is a z value by its use. This mark is only available in part (b).

2nd M1 for $\left(\frac{d-30}{5}\right)$ = their negative z value (or equivalent)

(c) M1 for a full method to find e. If they used $z = 1.2$ in (b) they can get M1 for $z = \pm$ 1.2 here
If they use symmetry about the mean $\mu + (\mu$ - their $d)$ then ft their d for M1
Must explicitly see the method used unless the answer is correct.

(d) M1 for a complete method or use of a correct expression e.g. "their 0.8849" - 0.1151
or **If their *d* < their *e*** using their values with $P(X < e) - P(X < d)$
If their $d \geq$ their e then they can only score from an argument like 1 – 2x0.1151
A negative probability or probability > 1 for part (d) scores M0A0