Centre No.					Paper Reference	Surname	Initial(s)
Candidate No.					6 6 8 3 / 0 1	Signature	

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 20 May 2011 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
8	
Total	

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 or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
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 in this question paper. The total mark for this paper is 75.
There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
P38164A
W850/R6683/57570 3/3/3

P 3 8 1 6 4 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

Leave blank

1. On a particular day the height above sea level, x metres, and the mid-day temperature, y°C, were recorded in 8 north European towns. These data are summarised below

$$S_{xx} = 3\ 535\ 237.5 \quad \sum y = 181 \quad \sum y^2 = 4305 \quad S_{xy} = -23\ 726.25$$

(a) Find S_{yy}

(2)

(b) Calculate, to 3 significant figures, the product moment correlation coefficient for these data.

(2)

(c) Give an interpretation of your coefficient.

(1)

A student thought that the calculations would be simpler if the height above sea level, h, was measured in kilometres and used the variable $h = \frac{x}{1000}$ instead of x.

(d) Write down the value of S_{hh}

(1)

(e) Write down the value of the correlation coefficient between h and y.

(1)

a) $S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$

$= 4305 - \frac{(181)^2}{8}$

$= 209.875$

b) $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

$r = \frac{-23726.25}{\sqrt{(3535237.5)(209.875)}}$

$= -0.871$ (3sf)

c) As the sea level increases the mid-day temperature decreases.

Question 1 continued

d/ 3.5352375

e/ −0.871 (3sf)

Leave blank

Q1

(Total 7 marks)

Leave blank

2. The random variable $X \sim \mathrm{N}(\mu, 5^2)$ and $\mathrm{P}(X < 23) = 0.9192$

(a) Find the value of μ.

(4)

(b) Write down the value of $\mathrm{P}(\mu < X < 23)$.

(1)

a) $z = \frac{X - \mu}{\sigma}$

$p = 0.9192 \quad \therefore \quad z = 1.40$ (tables)

$1.40 = \frac{23 - \mu}{5}$

$7 = 23 - \mu$

$\mu = 16$

b) 0.4192 (μ is at 50%)

Leave blank

3. The discrete random variable Y has probability distribution

3Y+2 5 8 11 14

y	1	2	3	4
$\mathrm{P}(Y = y)$	a	b	0.3	c

0.1 0.4 0.2 = 1

where a, b and c are constants.

The cumulative distribution function $\mathrm{F}(y)$ of Y is given in the following table

y	1	2	3	4
$\mathrm{F}(y)$	0.1	0.5	d	1.0

+0.4 0.8 +0.3 +0.2

where d is a constant.

(a) Find the value of a, the value of b, the value of c and the value of d.

(5)

(b) Find $\mathrm{P}(3Y + 2 \geqslant 8)$.

(2)

a)
$a = 0.1$
$b = 0.4$
$c = 0.2$
$d = 0.8$

b/ $P(2, 3, 4) = \underline{\underline{0.9}}$

Leave blank

4. Past records show that the times, in seconds, taken to run 100 m by children at a school can be modelled by a normal distribution with a mean of 16.12 and a standard deviation of 1.60

A child from the school is selected at random.

(a) Find the probability that this child runs 100 m in less than 15 s.

(3)

On sports day the school awards certificates to the fastest 30% of the children in the 100 m race.

(b) Estimate, to 2 decimal places, the slowest time taken to run 100 m for which a child will be awarded a certificate.

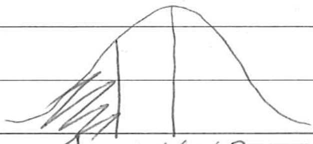
(4)

a) $Z = \dfrac{X - \mu}{\sigma}$

$= \dfrac{15 - 16.12}{1.6}$

$= -0.7$

$Z = -0.7 \quad \therefore \quad P = 1 - 0.7580$

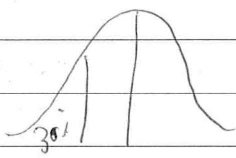


15 16.12

small bit $\therefore \quad P = 1 - 0.7580$

$= \underline{\underline{0.242}}$

b/



$P = 0.7 \quad \therefore \quad Z = -0.52$

$-$ (left of graph)

$-0.52 = \dfrac{X - 16.12}{1.6}$

$X = 15.288$ seconds

$= 15.29$ seconds (2dp)

Leave blank

5. A class of students had a sudoku competition. The time taken for each student to complete the sudoku was recorded to the nearest minute and the results are summarised in the table below.

Time	Mid-point, x	Frequency, f	
2 - 8	5	2	
9 - 12	10.5	7	9
13 - 15	14	5	14
15.5 16 - 18 18.5	17	8	22
19 - 22	20.5	4	26
23 - 30	26.5	4	30

(You may use $\sum fx^2 = 8603.75$)

(a) Write down the mid-point for the 9 - 12 interval. **(1)**

(b) Use linear interpolation to estimate the median time taken by the students. **(2)**

(c) Estimate the mean and standard deviation of the times taken by the students. **(5)**

The teacher suggested that a normal distribution could be used to model the times taken by the students to complete the sudoku.

(d) Give a reason to support the use of a normal distribution in this case. **(1)**

On another occasion the teacher calculated the quartiles for the times taken by the students to complete a different sudoku and found

$$Q_1 = 8.5 \qquad Q_2 = 13.0 \qquad Q_3 = 21.0$$

(e) Describe, giving a reason, the skewness of the times on this occasion. **(2)**

b/ $15.5 + \frac{1}{8}(3)$

$= 15.875$

c/ mean $= \frac{\Sigma fx}{n} = \frac{477.5}{30} = 15.916$

Leave blank

Question 5 continued

~~df~~ $\sigma = \sqrt{\frac{\Sigma f x^2}{n} - \left(\frac{\Sigma f x}{n}\right)^2}$

$= \sqrt{\frac{8603.75}{30} - (15.91\dot{6})^2}$

$= 5.783717566$

$= 5.784$ (3dp)

d) The mean is almost equal to the median

e) The median is closer to the lower quartile.
This is a positive skew.

Leave blank

6. Jake and Kamil are sometimes late for school.
The events J and K are defined as follows

J = the event that Jake is late for school
K = the event that Kamil is late for school

$\mathrm{P}(J) = 0.25,\ \mathrm{P}(J \cap K) = 0.15$ and $\mathrm{P}(J' \cap K') = 0.7$

On a randomly selected day, find the probability that

(a) at least one of Jake or Kamil are late for school, **(1)**

(b) Kamil is late for school. **(2)**

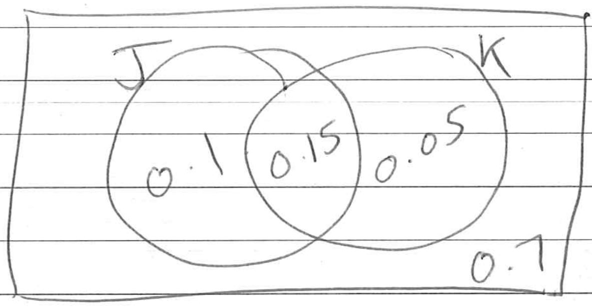
Given that Jake is late for school,

(c) find the probability that Kamil is late. **(3)**

The teacher suspects that Jake being late for school and Kamil being late for school are linked in some way.

(d) Determine whether or not J and K are statistically independent. **(2)**

(e) Comment on the teacher's suspicion in the light of your calculation in (d). **(1)**



a) 0.3

b) 0.2

c) $\dfrac{0.15}{0.25} = 0.6$

Leave blank

Question 6 continued

d/ if independent: $P(J) \times P(K) = P(J \cap K)$

$0.25 \times 0.2 = 0.05$

$P(J \cap K) \neq 0.05$

J and K are not statistically independent

e/ it appears that there is a link between Jake and Kamil being late for school.

Turn over

Leave blank

7. A teacher took a random sample of 8 children from a class. For each child the teacher recorded the length of their left foot, f cm, and their height, h cm. The results are given in the table below.

f	23	26	23	22	27	24	20	21
h	135	144	134	136	140	134	130	132

(You may use $\sum f = 186$ $\sum h = 1085$ $S_{ff} = 39.5$ $S_{hh} = 139.875$ $\sum fh = 25291$)

(a) Calculate S_{fh} **(2)**

(b) Find the equation of the regression line of h on f in the form $h = a + bf$.
Give the value of a and the value of b correct to 3 significant figures. **(5)**

(c) Use your equation to estimate the height of a child with a left foot length of 25 cm. **(2)**

(d) Comment on the reliability of your estimate in (c), giving a reason for your answer. **(2)**

The left foot length of the teacher is 25 cm.

(e) Give a reason why the equation in (b) should not be used to estimate the teacher's height. **(1)**

$S_{fh} = \sum fh - \frac{(\sum f)(\sum h)}{n}$

$= 25291 - \frac{(186)(1085)}{8}$

$= 64.75$

b/ $b = \frac{S_{fh}}{S_{ff}}$ $a = \bar{h} - b\bar{f}$

$= \frac{64.75}{39.5}$

$= 1.639240506$

Leave blank

Question 7 continued

$\bar{h} = \frac{1085}{8}$

$= 135.625$

$\bar{f} = \frac{186}{8}$

$= 23.25$

$a = 135.625 - 1.639240506(23.25)$

$a = 97.51265823$

$h = 97.5 + 1.64f$

c) $f = 25$

$h = 97.5 + 1.64(25)$

$= \underline{\underline{138.5\text{cm}}}$

d) ~~3~~ 25 is within the range provided so it is a reliable estimate

e) It should not because the data provided was for children

Leave blank

8. A spinner is designed so that the score S is given by the following probability distribution.

s	0	1	2	4	5
$P(S = s)$	p	0.25	0.25	0.20	0.20

0·1

(a) Find the value of p.
(2)

(b) Find $E(S)$.
(2)

(c) Show that $E(S^2) = 9.45$
(2)

(d) Find $Var(S)$.
(2)

Tom and Jess play a game with this spinner. The spinner is spun repeatedly and S counters are awarded on the outcome of each spin. If S is even then Tom receives the counters and if S is odd then Jess receives them. The first player to collect 10 or more counters is the winner.

(e) Find the probability that Jess wins after 2 spins.
(2)

(f) Find the probability that Tom wins after exactly 3 spins.
(4)

(g) Find the probability that Jess wins after exactly 3 spins.
(3)

a) $p = 0.1$

b) $0 \times 0.1 + 1 \times 0.25 + 2 \times 0.25 \times 4 \times 0.2 \times 5 \times 0.2$

$= 2.55$

c) $E(S^2) = 0^2 \times 0.1 + 1^2 \times 0.25 + 2^2 \times 0.25 \times 4^2 \times 0.2 + 5^2 \times 0.2$

$= 9.45$

d) $Var(S) = 9.45 - 2.55^2$

$= 2.9475$

Leave blank

Question 8 continued

e/ $P(5,5) = 0.2 \times 0.2$

$= \underline{0.04}$

f/ $P(4,4,2) = 0.2 \times 0.2 \times 0.25 = \frac{1}{100}$

$P(4,2,4) = \frac{1}{100}$

$P(2,4,4) = \frac{1}{100}$

$P(4,4,4) = 0.2 \times 0.2 \times 0.2 = 0.008$

$= 0.03 + 0.008 = \underline{0.038}$

g/ $P(5,5,5') = 0.2 \times 0.2 \times 0.8 = \frac{4}{125}$

$P(5,5',5) = \frac{4}{125}$

$P(5',5,5) = 4/125$

$= \underline{\frac{12}{125}}$