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

Paper Reference(s)

6683/01R

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 17 May 2013 – Morning

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	
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 7 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.
P43956A
W850/R6638/57570 5/5/
P 4 3 9 5 6 A 0 1 2 4

Turn over

PEARSON

Leave blank

1. Sammy is studying the number of units of gas, g, and the number of units of electricity, e, used in her house each week. A random sample of 10 weeks use was recorded and the data for each week were coded so that $x = \dfrac{g - 60}{4}$ and $y = \dfrac{e}{10}$. The results for the coded data are summarised below

$$\sum x = 48.0 \quad \sum y = 58.0 \quad S_{xx} = 312.1 \quad S_{yy} = 2.10 \quad S_{xy} = 18.35$$

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

Give the values of a and b correct to 3 significant figures.

(4)

(b) Hence find the equation of the regression line of e on g in the form $e = c + dg$.

Give the values of c and d correct to 2 significant figures.

(4)

(c) Use your regression equation to estimate the number of units of electricity used in a week when 100 units of gas were used.

(2)

Leave blank

Question 1 continued

Leave blank

Question 1 continued

Q1

(Total 10 marks)

P 4 3 9 5 6 A 0 4 2 4

2. The discrete random variable X takes the values 1, 2 and 3 and has cumulative distribution function $F(x)$ given by

x	1	2	3
$F(x)$	0.4	0.65	1

(a) Find the probability distribution of X.

(3)

(b) Write down the value of $F(1.8)$.

(1)

Q2

(Total 4 marks)

Leave blank

3. An agriculturalist is studying the yields, y kg, from tomato plants. The data from a random sample of 70 tomato plants are summarised below.

Yield (y kg)	**Frequency** (f)	**Yield midpoint** (x kg)
$0 \leqslant y < 5$	16	2.5
$5 \leqslant y < 10$	24	7.5
$10 \leqslant y < 15$	14	12.5
$15 \leqslant y < 25$	12	20
$25 \leqslant y < 35$	4	30

(You may use $\sum \text{f}x = 755$ and $\sum \text{f}x^2 = 12\,037.5$)

A histogram has been drawn to represent these data.

The bar representing the yield $5 \leqslant y < 10$ has a width of 1.5 cm and a height of 8 cm.

(a) Calculate the width and the height of the bar representing the yield $15 \leqslant y < 25$
(3)

(b) Use linear interpolation to estimate the median yield of the tomato plants.
(2)

(c) Estimate the mean and the standard deviation of the yields of the tomato plants.
(4)

(d) Describe, giving a reason, the skewness of the data.
(2)

(e) Estimate the number of tomato plants in the sample that have a yield of more than 1 standard deviation above the mean.
(2)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

P 4 3 9 5 6 A 0 8 2 4

Leave blank

Question 3 continued

Q3

(Total 13 marks)

P 4 3 9 5 6 A 0 9 2 4

Turn over

Leave blank

4. The time, in minutes, taken to fly from London to Malaga has a normal distribution with mean 150 minutes and standard deviation 10 minutes.

(a) Find the probability that the next flight from London to Malaga takes less than 145 minutes.

(3)

The time taken to fly from London to Berlin has a normal distribution with mean 100 minutes and standard deviation *d* minutes.

Given that 15% of the flights from London to Berlin take longer than 115 minutes,

(b) find the value of the standard deviation *d*.

(4)

The time, X minutes, taken to fly from London to another city has a normal distribution with mean μ minutes.

Given that $P(X < \mu - 15) = 0.35$

(c) find $P(X > \mu + 15 \mid X > \mu - 15)$.

(3)

Leave blank

Question 4 continued

P 4 3 9 5 6 A 0 1 1 2 4

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 10 marks)

P 4 3 9 5 6 A 0 1 3 2 4

Leave blank

5. A researcher believes that parents with a short family name tended to give their children a long first name. A random sample of 10 children was selected and the number of letters in their family name, x, and the number of letters in their first name, y, were recorded.

The data are summarised as:

$$\sum x = 60, \quad \sum y = 61, \quad \sum y^2 = 393, \quad \sum xy = 382, \quad S_{xx} = 28$$

(a) Find S_{yy} and S_{xy}

(3)

(b) Calculate the product moment correlation coefficient, r, between x and y.

(2)

(c) State, giving a reason, whether or not these data support the researcher's belief.

(2)

The researcher decides to add a child with family name "Turner" to the sample.

(d) Using the definition $S_{xx} = \sum (x - \bar{x})^2$, state the new value of S_{xx} giving a reason for your answer.

(2)

Given that the addition of the child with family name "Turner" to the sample leads to an increase in S_{yy}

(e) use the definition $S_{xy} = \sum (x - \bar{x})(y - \bar{y})$ to determine whether or not the value of r will increase, decrease or stay the same. Give a reason for your answer.

(2)

P 4 3 9 5 6 A 0 1 4 2 4

Leave blank

Question 5 continued

P 4 3 9 5 6 A 0 1 5 2 4

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Q5

(Total 11 marks)

P 4 3 9 5 6 A 0 1 7 2 4

Leave blank

6.

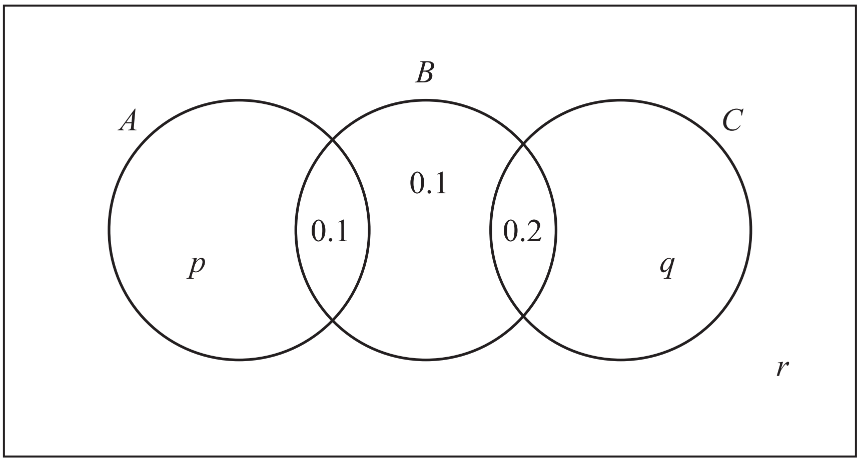

Figure 1

The Venn diagram in Figure 1 shows three events A, B and C and the probabilities associated with each region of B. The constants p, q and r each represent probabilities associated with the three separate regions outside B.

The events A and B are independent.

(a) Find the value of p.

(3)

Given that $\mathrm{P}(B \mid C) = \frac{5}{11}$

(b) find the value of q and the value of r.

(4)

(c) Find $\mathrm{P}(A \cup C \mid B)$.

(2)

P 4 3 9 5 6 A 0 1 8 2 4

Leave blank

Question 6 continued

P 4 3 9 5 6 A 0 1 9 2 4

Leave blank

Question 6 continued

Leave blank

Question 6 continued

(Total 9 marks)

Q6

P 4 3 9 5 6 A 0 2 1 2 4

Turn over

Leave blank

7. The score S when a spinner is spun has the following probability distribution.

s	0	1	2	4	5
$P(S = s)$	0.2	0.2	0.1	0.3	0.2

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

(b) Show that $E(S^2) = 10.4$
(2)

(c) Hence find $Var(S)$.
(2)

(d) Find

(i) $E(5S - 3)$,

(ii) $Var(5S - 3)$.
(4)

(e) Find $P(5S - 3 > S + 3)$
(3)

The spinner is spun twice.

The score from the first spin is S_1 and the score from the second spin is S_2

The random variables S_1 and S_2 are independent and the random variable $X = S_1 \times S_2$

(f) Show that $P(\{S_1 = 1\} \cap X < 5) = 0.16$
(2)

(g) Find $P(X < 5)$.
(3)

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Q7

(Total 18 marks)

TOTAL FOR PAPER: 75 MARKS

END

P 4 3 9 5 6 A 0 2 4 2 4