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

Thursday 15 May 2008 – 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
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 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 for 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 28 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.

This publication may be reproduced only in accordance with Edexcel Limited copyright policy.
©2008 Edexcel Limited.

Printer's Log. No.
H32582A
W850/R6683/57570 3/3/3/2/

H 3 2 5 8 2 A 0 1 2 8

Turn over

edexcel
advancing learning, changing lives

Leave blank

1. A disease is known to be present in 2% of a population. A test is developed to help determine whether or not someone has the disease.

Given that a person has the disease, the test is positive with probability 0.95

Given that a person does not have the disease, the test is positive with probability 0.03

(a) Draw a tree diagram to represent this information.

(3)

A person is selected at random from the population and tested for this disease.

(b) Find the probability that the test is positive.

(3)

A doctor randomly selects a person from the population and tests him for the disease. Given that the test is positive,

(c) find the probability that he does not have the disease.

(2)

(d) Comment on the usefulness of this test.

(1)

Leave blank

Question 1 continued

Leave blank

2. The age in years of the residents of two hotels are shown in the back to back stem and leaf diagram below.

Abbey Hotel 8|5|0 means 58 years in Abbey hotel and 50 years in Balmoral hotel Balmoral Hotel

(1)	2	0		
(4)	9751	1		
(4)	9831	2	6	(1)
(11)	99997665332	3	447	(3)
(6)	987750	4	005569	(6)
(1)	8	5	000013667	(9)
		6	233457	(6)
		7	015	(3)

For the Balmoral Hotel,

(a) write down the mode of the age of the residents,

(1)

(b) find the values of the lower quartile, the median and the upper quartile.

(3)

(c) (i) Find the mean, $\bar{x}$, of the age of the residents.

(ii) Given that $\sum x^2 = 81\,213$ find the standard deviation of the age of the residents.

(4)

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

(d) Evaluate this measure for the Balmoral Hotel.

(2)

For the Abbey Hotel, the mode is 39, the mean is 33.2, the standard deviation is 12.7 and the measure of skewness is – 0.454

(e) Compare the two age distributions of the residents of each hotel.

(3)

Leave blank

Question 2 continued

Leave blank

3. 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) $\text{Var}(X)$,

(4)

(c) $E(2X - 4)$.

(2)

H 3 2 5 8 2 A 0 1 0 2 8

Leave blank

Question 3 continued

Leave blank

4. Crickets make a noise. The pitch, v kHz, of the noise made by a cricket was recorded at 15 different temperatures, t °C. These data are summarised below.

$$\sum t^2 = 10\,922.81, \sum v^2 = 42.3356, \sum tv = 677.971, \sum t = 401.3, \sum v = 25.08$$

(a) Find S_{tt}, S_{vv} and S_{tv} for these data.

(4)

(b) Find the product moment correlation coefficient between t and v.

(3)

(c) State, with a reason, which variable is the explanatory variable.

(2)

(d) Give a reason to support fitting a regression model of the form $v = a + bt$ to these data.

(1)

(e) Find the value of a and the value of b. Give your answers to 3 significant figures.

(4)

(f) Using this model, predict the pitch of the noise at 19 °C.

(1)

Leave blank

Question 4 continued

Leave blank

5. A person's blood group is determined by whether or not it contains any of 3 substances *A*, *B* and *C*.

A doctor surveyed 300 patients' blood and produced the table below.

Blood contains	**No. of Patients**
only *C*	100
A and *C* but not *B*	100
only A	30
B and *C* but not *A*	25
only *B*	12
A, *B* and *C*	10
A and *B* but not *C*	3

(a) Draw a Venn diagram to represent this information.

(4)

Leave blank

(b) Find the probability that a randomly chosen patient's blood contains substance *C*.
(2)

Harry is one of the patients. Given that his blood contains substance *A*,

(c) find the probability that his blood contains all 3 substances.
(2)

Patients whose blood contains none of these substances are called universal blood donors.

(d) Find the probability that a randomly chosen patient is a universal blood donor.
(2)

Leave blank

6. The discrete random variable X can take only the values 2, 3 or 4. For these values the cumulative distribution function is defined by

$$\mathrm{F}(x) = \frac{(x+k)^2}{25} \quad \text{for} \quad x = 2, 3, 4$$

where k is a positive integer.

(a) Find k.

(2)

(b) Find the probability distribution of X.

(3)

H 3 2 5 8 2 A 0 2 2 2 8

Leave blank

7. A packing plant fills bags with cement. The weight X kg of a bag of cement can be modelled by a normal distribution with mean 50 kg and standard deviation 2 kg.

(a) Find $P(X > 53)$.

(3)

(b) Find the weight that is exceeded by 99% of the bags.

(5)

Three bags are selected at random.

(c) Find the probability that two weigh more than 53 kg and one weighs less than 53 kg.

(4)

Leave blank

Question 7 continued

Q7

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END

H 3 2 5 8 2 A 0 2 7 2 8