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 27 May 2010 – 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 (Pink)

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 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 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.
©2010 Edexcel Limited.

Printer's Log. No.
H35395A
W850/R6683/57570 4/5/4/4

H 3 5 3 9 5 A 0 1 2 8

Turn over

edexcel
advancing learning, changing lives

Leave blank

1. Gary compared the total attendance, x, at home matches and the total number of goals, y, scored at home during a season for each of 12 football teams playing in a league. He correctly calculated:

$$S_{xx} = 1022500 \quad S_{yy} = 130.9 \quad S_{xy} = 8825$$

(a) Calculate the product moment correlation coefficient for these data.

(2)

(b) Interpret the value of the correlation coefficient.

(1)

Helen was given the same data to analyse. In view of the large numbers involved she decided to divide the attendance figures by 100. She then calculated the product moment correlation coefficient between $\frac{x}{100}$ and y.

(c) Write down the value Helen should have obtained.

(1)

H 3 5 3 9 5 A 0 2 2 8

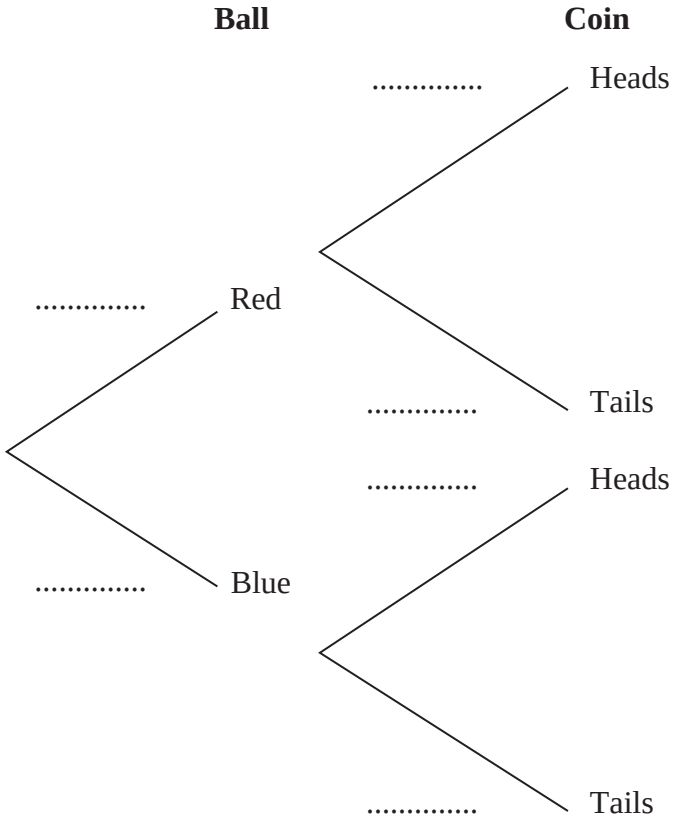
Leave blank

2. An experiment consists of selecting a ball from a bag and spinning a coin. The bag contains 5 red balls and 7 blue balls. A ball is selected at random from the bag, its colour is noted and then the ball is returned to the bag.

When a red ball is selected, a biased coin with probability $\frac{2}{3}$ of landing heads is spun.

When a blue ball is selected a fair coin is spun.

(a) Complete the tree diagram below to show the possible outcomes and associated probabilities.



(2)

Shivani selects a ball and spins the appropriate coin.

(b) Find the probability that she obtains a head.

(2)

Given that Tom selected a ball at random and obtained a head when he spun the appropriate coin,

(c) find the probability that Tom selected a red ball.

(3)

Shivani and Tom each repeat this experiment.

(d) Find the probability that the colour of the ball Shivani selects is the same as the colour of the ball Tom selects.

(3)

Leave blank

Question 2 continued

Leave blank

3. The discrete random variable X has probability distribution given by

x	-1	0	1	2	3
$\mathrm{P}(X = x)$	$\frac{1}{5}$	a	$\frac{1}{10}$	a	$\frac{1}{5}$

where a is a constant.

(a) Find the value of a.

(2)

(b) Write down $\mathrm{E}(X)$.

(1)

(c) Find $\mathrm{Var}(X)$.

(3)

The random variable $Y = 6 - 2X$

(d) Find $\mathrm{Var}(Y)$.

(2)

(e) Calculate $\mathrm{P}(X \geqslant Y)$.

(3)

Leave blank

Question 3 continued

Leave blank

4. The Venn diagram in Figure 1 shows the number of students in a class who read any of 3 popular magazines A, B and C.

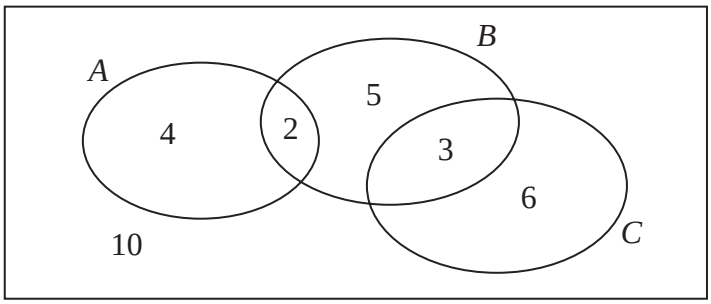


Figure 1

One of these students is selected at random.

(a) Show that the probability that the student reads more than one magazine is $\frac{1}{6}$.
(2)

(b) Find the probability that the student reads A or B (or both).
(2)

(c) Write down the probability that the student reads both A and C.
(1)

Given that the student reads at least one of the magazines,

(d) find the probability that the student reads C.
(2)

(e) Determine whether or not reading magazine B and reading magazine C are statistically independent.
(3)

H 3 5 3 9 5 A 0 1 2 2 8

Leave blank

Question 4 continued

Leave blank

5. A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

Hours	1 – 10	11 – 20	21 – 25	26 – 30	31 – 40	41 – 59
Frequency	6	15	11	13	8	3
Mid-point	5.5	15.5		28		50

(a) Find the mid-points of the $21-25$ hour and $31-40$ hour groups.

(2)

A histogram was drawn to represent these data. The $11-20$ group was represented by a bar of width 4 cm and height 6 cm.

(b) Find the width and height of the $26-30$ group.

(3)

(c) Estimate the mean and standard deviation of the time spent watching television by these students.

(5)

(d) Use linear interpolation to estimate the median length of time spent watching television by these students.

(2)

The teacher estimated the lower quartile and the upper quartile of the time spent watching television to be 15.8 and 29.3 respectively.

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

(2)

Leave blank

Question 5 continued

Leave blank

6. A travel agent sells flights to different destinations from *Beerow* airport. The distance d, measured in 100 km, of the destination from the airport and the fare £f are recorded for a random sample of 6 destinations.

Destination	A	B	C	D	E	F
d	2.2	4.0	6.0	2.5	8.0	5.0
f	18	20	25	23	32	28

[You may use $\sum d^2 = 152.09$ $\sum f^2 = 3686$ $\sum fd = 723.1$]

(a) Using the axes below, complete a scatter diagram to illustrate this information.
(2)

(b) Explain why a linear regression model may be appropriate to describe the relationship between f and d.
(1)

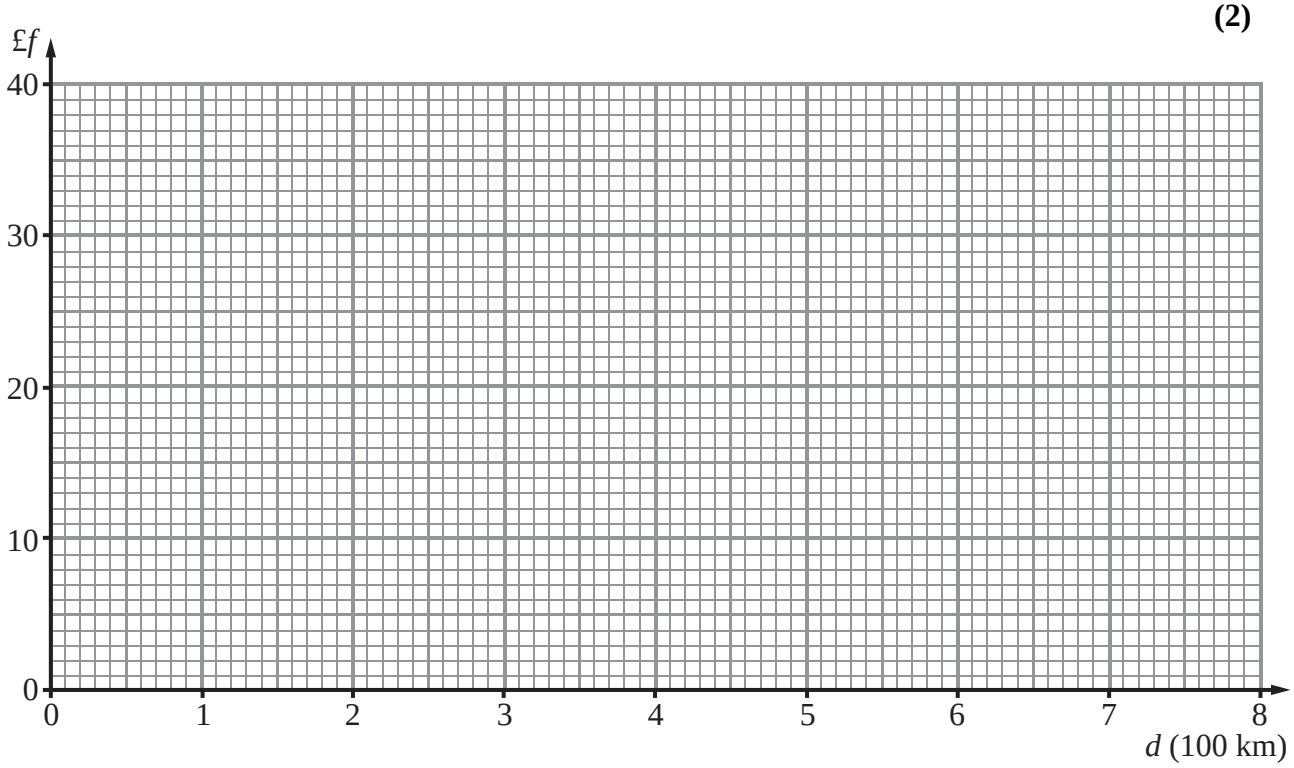
(c) Calculate S_{dd} and S_{fd}
(4)

(d) Calculate the equation of the regression line of f on d giving your answer in the form $f = a + bd$.
(4)

(e) Give an interpretation of the value of b.
(1)

Jane is planning her holiday and wishes to fly from *Beerow* airport to a destination t km away. A rival travel agent charges 5p per km.

(f) Find the range of values of t for which the first travel agent is cheaper than the rival.
(2)



H 3 5 3 9 5 A 0 2 0 2 8

Leave blank

Question 6 continued

Leave blank

7. The distances travelled to work, D km, by the employees at a large company are normally distributed with $D \sim \mathrm{N}(30, 8^2)$.

(a) Find the probability that a randomly selected employee has a journey to work of more than 20 km.

(3)

(b) Find the upper quartile, Q_3, of D.

(3)

(c) Write down the lower quartile, Q_1, of D.

(1)

An outlier is defined as any value of D such that $D < h$ or $D > k$ where

$$h = Q_1 - 1.5 \times (Q_3 - Q_1) \quad \text{and} \quad k = Q_3 + 1.5 \times (Q_3 - Q_1)$$

(d) Find the value of h and the value of k.

(2)

An employee is selected at random.

(e) Find the probability that the distance travelled to work by this employee is an outlier.

(3)

Leave blank

Question 7 continued

Q7

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END