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

Paper Reference(s)

6686/01

Edexcel GCE

Statistics S4

Advanced/Advanced Subsidiary

Friday 21 June 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	
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 6 questions in this question paper. The total mark for this paper is 75.
There are 20 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.
P41806A
W850/R6686/57570 5/5/5/5/4/5/5/

P 4 1 8 0 6 A 0 1 2 0

Turn over

PEARSON

Leave blank

1. George owns a garage and he records the mileage of cars, x thousands of miles, between services. The results from a random sample of 10 cars are summarised below.

$$\sum x = 113.4 \qquad \sum x^2 = 1414.08$$

The mileage of cars between services is normally distributed and George believes that the standard deviation is 2.4 thousand miles.

Stating your hypotheses clearly, test, at the 5% level of significance, whether or not these data support George's belief.

(7)

Leave blank

2. Every 6 months some engineers are tested to see if their times, in minutes, to assemble a particular component have changed. The times taken to assemble the component are normally distributed. A random sample of 8 engineers was chosen and their times to assemble the component were recorded in January and in July. The data are given in the table below.

Engineer	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
January	17	19	22	26	15	28	18	21
July	19	18	25	24	17	25	16	19

(a) Calculate a 95% confidence interval for the mean difference in times.

(7)

(b) Use your confidence interval to state, giving a reason, whether or not there is evidence of a change in the mean time to assemble a component. State your hypotheses clearly.

(3)

Leave blank

3. An archaeologist is studying the compression strength of bricks at some ancient European sites. He took random samples from two sites A and B and recorded the compression strength of these bricks in appropriate units. The results are summarised below.

Site	Sample size (n)	Sample mean ($\bar{x}$)	Standard deviation (s)
A	7	8.43	4.24
B	13	14.31	4.37

It can be assumed that the compression strength of bricks is normally distributed.

(a) Test, at the 2% level of significance, whether or not there is evidence of a difference in the variances of compression strength of the bricks between these two sites. State your hypotheses clearly.
(5)

Site A is older than site B and the archaeologist claims that the mean compression strength of the bricks was greater at the younger site.

(b) Stating your hypotheses clearly and using a 1% level of significance, test the archaeologist's claim.
(6)

(c) Explain briefly the importance of the test in part (a) to the test in part (b).
(1)

Leave blank

Question 3 continued

Leave blank

4. A random sample of size 2, X_1 and X_2, is taken from the random variable X which has a continuous uniform distribution over the interval $[-a, 2a]$, $a > 0$

(a) Show that $\overline{X} = \dfrac{X_1 + X_2}{2}$ is a biased estimator of a and find the bias. **(3)**

The random variable $Y = k\overline{X}$ is an unbiased estimator of a.

(b) Write down the value of the constant k. **(1)**

(c) Find Var(Y). **(4)**

The random variable M is the maximum of X_1 and X_2

The probability density function, $m(x)$, of M is given by

$$m(x) = \begin{cases} \dfrac{2(x+a)}{9a^2} & -a \leqslant x \leqslant 2a \\ 0 & \text{otherwise} \end{cases}$$

(d) Show that M is an unbiased estimator of a. **(4)**

Given that $E(M^2) = \dfrac{3}{2}a^2$

(e) find Var(M). **(1)**

(f) State, giving a reason, whether you would use Y or M as an estimator of a. **(2)**

A random sample of two values of X are 5 and –1

(g) Use your answer to part (f) to estimate a. **(1)**

Leave blank

Question 4 continued

Leave blank

5. Water is tested at various stages during a purification process by an environmental scientist. A certain organism occurs randomly in the water at a rate of λ every 10 ml. The scientist selects a random sample of 20 ml of water to check whether there is evidence that λ is greater than 1. The criterion the scientist uses for rejecting the hypothesis that $\lambda = 1$ is that there are 4 or more organisms in the sample of 20 ml.

(a) Find the size of the test.

(2)

(b) When $\lambda = 2.5$ find P(Type II error).

(2)

A statistician suggests using an alternative test. The statistician's test involves taking a random sample of 10 ml and rejecting the hypothesis that $\lambda = 1$ if 2 or more organisms are present but accepting the hypothesis if no organisms are in the sample. If only 1 organism is found then a second random sample of 10 ml is taken and the hypothesis is rejected if 2 or more organisms are present, otherwise the hypothesis is accepted.

(c) Show that the power of the statistician's test is given by

$$1 - e^{-\lambda} - \lambda(1 + \lambda)e^{-2\lambda}$$

(4)

Table 1 below gives some values, to 2 decimal places, of the power function of the statistician's test.

λ	1.5	2	2.5	3	3.5	4
Power	0.59	0.75	0.86	r	0.96	0.97

Table 1

(d) Find the value of r.

(1)

Question 5 continues on page 16

Leave blank

Question 5 continued

Question 5 continues on the next page

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

Leave blank

Question 5 continued

For your convenience Table 1 is repeated here.

λ	1.5	2	2.5	3	3.5	4
Power	0.59	0.75	0.86	r	0.96	0.97

Table 1

Figure 1 shows a graph of the power function for the scientist's test.

(e) On the same axes draw the graph of the power function for the statistician's test.
(2)

Given that it takes 20 minutes to collect and test a 20 ml sample and 15 minutes to collect and test a 10 ml sample

(f) show that the expected time of the statistician's test is slower than the scientist's test for $\lambda e^{-\lambda} > \frac{1}{3}$
(4)

(g) By considering the times when $\lambda = 1$ and $\lambda = 2$ together with the power curves in part (e) suggest, giving a reason, which test you would use.
(2)

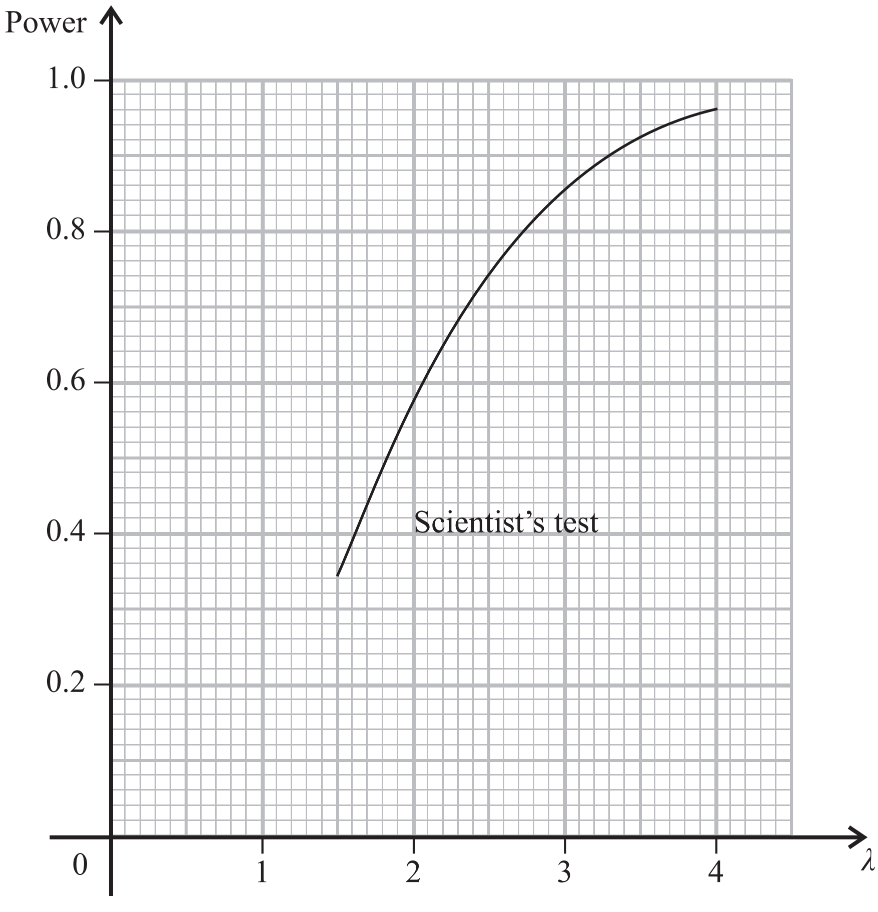

Figure 1

Leave blank

Question 5 continued

Q5

(Total 17 marks)

P 4 1 8 0 6 A 0 1 7 2 0

Leave blank

6. The carbon content, measured in suitable units, of steel is normally distributed. Two independent random samples of steel were taken from a refining plant at different times and their carbon content recorded. The results are given below.

Sample A:	1.5	0.9	1.3	1.2		
Sample B:	0.4	0.6	0.8	0.3	0.5	0.4

(a) Stating your hypotheses clearly, carry out a suitable test, at the 10% level of significance, to show that both samples can be assumed to have come from populations with a common variance σ^2.

(7)

(b) Showing your working clearly, find the 99% confidence interval for σ^2 based on both samples.

(6)

Leave blank

Question 6 continued

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