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

Paper Reference(s)

6686/01R

Edexcel GCE

Statistics S4

Advanced/Advanced Subsidiary

Thursday 12 June 2014 – 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	
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 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 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.
P43154A
W850/R6686/57570 5/5/5/5/

P 4 3 1 5 4 A 0 1 2 4

Turn over

PEARSON

Leave blank

1. In a trial for a new cough medicine, a random sample of 8 healthy patients were given steadily increasing doses of a pepper extract until they started coughing. The level of pepper that triggered the coughing was recorded. Each patient completed the trial after taking a standard cough medicine and, at a later time, after taking the new medicine. The results are given in the table below.

	Level of pepper extract that triggers coughing							
Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Standard medicine	46	12	18	31	23	16	27	9
New medicine	53	16	13	49	11	34	38	22

(a) Using a suitable test, at the 5% level of significance, state whether or not, on the basis of this trial, you would recommend using the new medicine. State your hypotheses clearly.

(8)

(b) State an assumption needed to carry out this test.

(1)

Leave blank

2. The cloth produced by a certain manufacturer has defects that occur randomly at a constant rate of λ per square metre. If λ is thought to be greater than 1.5 then action has to be taken.

Using $H_0: \lambda = 1.5$ and $H_1: \lambda > 1.5$ a quality control officer takes a 4 m^2 sample of cloth and rejects H_0 if there are 11 or more defects. If there are 8 or fewer defects she accepts H_0. If there are 9 or 10 defects a second sample of 4 m^2 is taken and H_0 is rejected if there are 11 or more defects in this second sample, otherwise it is accepted.

(a) Find the size of this test.

(4)

(b) Find the power of this test when $\lambda = 2$

(3)

Leave blank

3. A farmer is investigating the milk yields of two breeds of cow. He takes a random sample of 9 cows of breed *A* and an independent random sample of 12 cows of breed *B*. For a 5 day period he measures the amount of milk, x gallons, produced by each cow. The results are summarised in the table below.

Breed	Sample size	Mean ($\bar{x}$)	Standard deviation (s_x)
A	9	6.23	2.98
B	12	7.13	2.33

The amount of milk produced by each cow can be assumed to follow a normal distribution.

(a) Use a two-tail test to show, at the 10% level of significance, that the variances of the yields of the two breeds can be assumed to be equal. State your hypotheses clearly.

(4)

(b) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is a difference in the mean yields of the two breeds of cow.

(7)

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

(1)

Leave blank

Question 3 continued

Leave blank

4. At the start of each academic year, a large college carries out a diagnostic test on a random sample of new students. Past experience has shown that the standard deviation of the scores on this test is 19.71

The admissions tutor claimed that the new students in 2013 would have more varied scores than usual. The scores for the students taking the test can be assumed to come from a normal distribution. A random sample of 10 new students was taken and the score x, for each student was recorded. The data are summarised as $\sum x = 619$ $\sum x^2 = 42397$

(a) Stating your hypotheses clearly, and using a 5% level of significance, test the admission tutor's claim.

(6)

The admissions tutor decides that in future he will use the same hypotheses but take a larger sample of size 30 and use a significance level of 1%.

(b) Use the tables to show that, to 3 decimal places, the critical region for S^2 is $S^2 > 664.281$

(3)

(c) Find the probability of a type II error using this test when the true value of the standard deviation is in fact 22.20

(3)

Leave blank

Question 4 continued

Leave blank

5. A large company has designed an aptitude test for new recruits. The score, S, for an individual taking the test, has a normal distribution with mean μ and standard deviation σ.

In order to estimate μ and σ, a random sample of 15 new recruits were given the test and their scores, x, are summarised as

$$\sum x = 880 \quad \sum x^2 = 54\,892$$

(a) Calculate a 95% confidence interval for

(i) μ,

(ii) σ.

(11)

The company wants to ensure that no more than 80% of new recruits pass the test.

(b) Using values from your confidence intervals in part (a), estimate the lowest pass mark they should set.

(5)

Leave blank

Question 5 continued

Leave blank

6. Emily is monitoring the level of pollution in a river. Over a period of time she has found that the amount of pollution, X, in a 100 ml sample of river water has a continuous distribution with probability density function $f(x)$ given by

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

where a is a constant.

Emily takes a random sample $X_1, X_2, X_3, ..., X_n$ to try to estimate the value of a.

(a) Show that $E(\bar{X}) = \dfrac{2a}{3}$ and $\text{Var}(\bar{X}) = \dfrac{a^2}{18n}$

(4)

The random variable $S = p\bar{X}$, where p is a constant, is an unbiased estimator of a.

(b) Write down the value of p and find $\text{Var}(S)$.

(2)

Felix suggests using the statistic $M = \max\{X_1, X_2, X_3, ..., X_n\}$ as an estimator of a.

He calculates $E(M) = \dfrac{2n}{2n+1}a$ and $\text{Var}(M) = \dfrac{n}{(n+1)(2n+1)^2}a^2$

(c) State, giving your reasons, whether or not M is a consistent estimator of a.

(3)

The random variable $T = qM$, where q is a constant, is an unbiased estimator of a.

(d) Write down, in terms of n, the value of q and find $\text{Var}(T)$.

(3)

(e) State, giving your reasons, which of S or T you would recommend Emily use as an estimator of a.

(3)

Emily took a sample of 5 values of X and obtained the following:

5.3 4.3 5.7 7.8 6.9

(f) Calculate the estimate of a using your recommended estimator from part (e).

(2)

(g) Find the standard error of your estimate, giving your answer to 2 decimal places.

(2)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 19 marks)

TOTAL FOR PAPER: 75 MARKS

END