Centre No.						Paper Reference							Surname	Initial(s)
Candidate No.						6	6	9	1	/	0	1 R	Signature	

Paper Reference(s)

6691/01R

Edexcel GCE

Statistics S3

Advanced/Advanced Subsidiary

Thursday 13 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	
7	
8	
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 8 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.

This publication may be reproduced only in accordance with Pearson Education Ltd copyright policy.
©2013 Pearson Education Ltd.

Printer's Log. No.
P42965A
W850/R6684/57570 5/5/5/

P 4 2 9 6 5 A 0 1 2 4

Turn over

PEARSON

Leave blank

1. A gym club has 400 members of which 300 are males.

Explain clearly how a stratified sample of size 60 could be taken.

(3)

Q1

(Total 3 marks)

Leave blank

2. A random sample of size n is to be taken from a population that is normally distributed with mean 40 and standard deviation 3. Find the minimum sample size such that the probability of the sample mean being greater than 42 is less than 5%.

(5)

Q2

(Total 5 marks)

Leave blank

3. The table below shows the population and the number of council employees for different towns and villages.

Town or village	Population	Number of council employees
A	211	10
B	356	2
C	1047	12
D	2463	21
E	4892	16
F	6479	25
G	6571	67
H	6573	45
I	9845	48
J	14 784	34

(a) Find, to 3 decimal places, Spearman's rank correlation coefficient between the population and the number of council employees.

(5)

(b) Use your value of Spearman's rank correlation coefficient to test for evidence of a positive correlation between the population and the number of council employees. Use a 2.5% significance level. State your hypotheses clearly.

(4)

It is suggested that a product moment correlation coefficient would be a more suitable calculation in this case. The product moment correlation coefficient for these data is 0.627 to 3 decimal places.

(c) Use the value of the product moment correlation coefficient to test for evidence of a positive correlation between the population and the number of council employees. Use a 2.5% significance level.

(2)

(d) Interpret and comment on your results from part (b) and part (c).

(2)

Leave blank

Question 3 continued

Leave blank

4. John thinks that a person's eye colour is related to their hair colour. He takes a random sample of 600 people and records their eye and hair colours. The results are shown in Table 1.

		Hair colour				
		Black	Brown	Red	Blonde	Total
Eye colour	Brown	45	125	15	58	243
	Blue	34	90	10	58	192
	Hazel	20	38	16	26	100
	Green	6	29	7	23	65
	Total	105	282	48	165	600

Table 1

John carries out a χ^2 test in order to test whether eye colour and hair colour are related. He calculates the expected frequencies shown in Table 2.

		Hair colour			
		Black	Brown	Red	Blonde
Eye colour	Brown	42.5	114.2	19.4	66.8
	Blue	33.6	90.2	15.4	52.8
	Hazel	17.5	47	8	27.5
	Green	11.4	30.6	5.2	17.9

Table 2

(a) Show how the value 47 in Table 2 has been calculated.

(1)

(b) Write down the number of degrees of freedom John should use in this χ^2 test.

(1)

Given that the value of the χ^2 statistic is 20.6, to 3 significant figures,

(c) find the smallest value of α for which the null hypothesis will be rejected at the α% level of significance.

(1)

(d) Use the data from Table 1 to test at the 5% level of significance whether or not the proportions of people in the population with black, brown, red and blonde hair are in the ratio 2:6:1:3
State your hypotheses clearly.

(9)

Leave blank

Question 4 continued

Leave blank

5. A manufacturer produces circular discs with diameter D mm, such that $D \sim \mathrm{N}(\mu, \sigma^2)$.
A random sample of discs is taken and, using tables of the normal distribution, a 90% confidence interval for μ is found to be

$$(118.8, 121.2)$$

(a) Find a 98% confidence interval for μ.

(6)

(b) Hence write down a 98% confidence interval for the circumference of the discs.

(1)

Using three different random samples, three 98% confidence intervals for μ are to be found.

(c) Calculate the probability that all the intervals will contain μ.

(2)

Leave blank

Question 5 continued

Leave blank

6. The continuous random variable X is uniformly distributed over the interval

$$[a - 1, a + 5]$$

where a is a constant.

Fifty observations of X are taken, giving a sample mean of 17.2

(a) Use the Central Limit Theorem to find an approximate distribution for $\bar{X}$.

(3)

(b) Hence find a 95% confidence interval for a.

(4)

Leave blank

7. A farmer monitored the amount of lead in soil in a field next to a factory.
He took 100 samples of soil, randomly selected from different parts of the field, and found the mean weight of lead to be 67 mg/kg with standard deviation 25 mg/kg.
After the factory closed, the farmer took 150 samples of soil, randomly selected from different parts of the field, and found the mean weight of lead to be 60 mg/kg with standard deviation 10 mg/kg.

(a) Test at the 5% level of significance whether or not the mean weight of lead in the soil decreased after the factory closed. State your hypotheses clearly.
(7)

(b) Explain the significance of the Central Limit Theorem to the test in part (a).
(1)

(c) State an assumption you have made to carry out this test.
(1)

Leave blank

8. A farmer supplies both duck eggs and chicken eggs. The weights of duck eggs, D grams, and chicken eggs, C grams, are such that

$$D \sim \mathrm{N}(54,\ 1.2^2) \text{ and } C \sim \mathrm{N}(44,\ 0.8^2).$$

(a) Find the probability that the weights of 2 randomly selected duck eggs will differ by more than 3 g.

(6)

(b) Find the probability that the weight of a randomly selected chicken egg is less than $\frac{4}{5}$ of the weight of a randomly selected duck egg.

(5)

Eggs are packed in boxes which contain either 6 randomly selected duck eggs or 6 randomly selected chicken eggs. The weight of an empty box has distribution $\mathrm{N}\left(28, \sqrt{5}^2\right)$.

(c) Find the probability that a full box of duck eggs weighs at least 50 g more than a full box of chicken eggs.

(6)

Leave blank

Question 8 continued