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 22 May 2014 – 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 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 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 Pearson Education Ltd copyright policy.
©2014 Pearson Education Ltd.

Printer's Log. No.
P43155A
W850/R6691/57570 5/5/5/5/

P 4 3 1 5 5 A 0 1 2 8

Turn over

PEARSON

Leave blank

1. A journalist is investigating factors which influence people when they buy a new car. One possible factor is fuel efficiency. The journalist randomly selects 8 car models. Each model's annual sales and fuel efficiency, in km/litre, are shown in the table below.

Car model	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Annual sales	1800	5400	18 100	7100	9300	4800	12 200	10 700
Fuel efficiency	5.2	18.6	14.8	13.2	18.3	11.9	16.5	17.7

(a) Calculate Spearman's rank correlation coefficient for these data.

(5)

The journalist believes that car models with higher fuel efficiency will achieve higher sales.

(b) Stating your hypotheses clearly, test whether or not the data support the journalist's belief. Use a 5% level of significance.

(4)

(c) State the assumption necessary for a product moment correlation coefficient to be valid in this case.

(1)

(d) The mean and median fuel efficiencies of the car models in the random sample are 14.5 km/litre and 15.65 km/litre respectively. Considering these statistics, as well as the distribution of the fuel efficiency data, state whether or not the data suggest that the assumption in part (c) might be true in this case. Give a reason for your answer. (No further calculations are required.)

(1)

Leave blank

Question 1 continued

Leave blank

Question 1 continued

Leave blank

Question 1 continued

(Total 11 marks)

Q1

Leave blank

2. A survey asked a random sample of 200 people their age and the main use of their mobile phone.

The results are shown in Table 1 below.

		Main use of their mobile phone		
		Internet	**Texts**	**Phone calls**
Age	**Under 20**	27	14	9
	From 20 to 40	32	34	29
	Over 40	15	19	21

Table 1

The data are to be used to test whether or not age and main use of their mobile phone are independent.

Table 2 shows the expected frequencies for each group, assuming people's age and main use of their mobile phone are independent.

		Main use of their mobile phone		
		Internet	**Texts**	**Phone calls**
Age	**Under 20**	18.5	16.75	14.75
	From 20 to 40	35.15	31.825	28.025
	Over 40	20.35	18.425	16.225

Table 2

(a) For users under 20 choosing the Internet as the main use of their mobile phone,

(i) verify that the expected frequency is 18.5

(ii) show that the contribution to the χ^2 test statistic is 3.91 to 3 significant figures.

(2)

(b) Given that the χ^2 test statistic for the data is 9.893 to 3 decimal places, test at the 5% level of significance whether or not age and main use of their mobile phone are independent. State your hypotheses clearly.

(5)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Q2

(Total 7 marks)

Leave blank

3. A company produces two types of milk powder, 'Semi-Skimmed' and 'Full Cream'. In tests, each type of milk powder is used to make a large number of cups of coffee. The mass, S grams, of 'Semi-Skimmed' milk powder used in one cup of coffee is modelled by $S \sim N(4.9, 0.8^2)$. The mass, C grams, of 'Full Cream' milk powder used in one cup of coffee is modelled by $C \sim N(2.5, 0.4^2)$

(a) Two cups of coffee, one with each type of milk powder, are to be selected at random. Find the probability that the mass of 'Semi-Skimmed' milk powder used will be at least double that of the 'Full Cream' milk powder used.

(6)

(b) 'Semi-Skimmed' milk powder is sold in 500 g packs. Find the probability that one pack will be sufficient for 100 cups of coffee.

(5)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Q3

(Total 11 marks)

Leave blank

4. A manufacturing company produces solar panels. The output of each solar panel is normally distributed with standard deviation 6 watts. It is thought that the mean output, μ, is 160 watts.

A researcher believes that the mean output of the solar panels is greater than 160 watts. He writes down the output values of 5 randomly selected solar panels. He uses the data to carry out a hypothesis test at the 5% level of significance.

He tests $H_0 : \mu = 160$ against $H_1 : \mu > 160$

On reporting to his manager, the researcher can only find 4 of the output values. These are shown below

168.2 157.4 173.3 161.1

Given that the result of the hypothesis test is that there is significant evidence to reject H_0 at the 5% level of significance, calculate the minimum possible missing output value, α. Give your answer correct to 1 decimal place.

(6)

P 4 3 1 5 5 A 0 1 4 2 8

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 6 marks)

Leave blank

5. A student believes that there is a difference in the mean lengths of English and French films. He goes to the university video library and randomly selects a sample of 120 English films and a sample of 70 French films. He notes the length, x minutes, of each of the films in his samples. His data are summarised in the table below.

	Σx	Σx^2	s^2	n
English films	10650	956909	98.5	120
French films	6510	615849	151	70

(a) Verify that the unbiased estimate of the variance, s^2, of the lengths of English films is 98.5 minutes^2

(2)

(b) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean lengths of English and French films are different.

(7)

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

(1)

(d) The university video library contained 724 English films and 473 French films. Explain how the student could have taken a stratified sample of 190 of these films.

(3)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Q5

(Total 13 marks)

Leave blank

6. Bags of £1 coins are paid into a bank. Each bag contains 20 coins.

The bank manager believes that 5% of the £1 coins paid into the bank are fakes. He decides to use the distribution $X \sim B(20, 0.05)$ to model the random variable X, the number of fake £1 coins in each bag.

(a) State the assumptions necessary for the binomial distribution to be an appropriate model in this case.

(2)

The bank manager checks a random sample of 150 bags of £1 coins and records the number of fake coins found in each bag. His results are summarised in Table 1.

Number of fake coins in each bag	0	1	2	3	4 or more
Observed frequency	43	62	26	13	6
Expected frequency	53.8	56.6	r	8.9	s

Table 1

(b) Calculate the values of r and s, giving your answers to 1 decimal place.

(3)

(c) Carry out a hypothesis test, at the 5% significance level, to see if the data supports the bank manager's statistical model. State your hypotheses clearly.

(7)

Question 6 parts (d) and (e) are continued on page 24

P 4 3 1 5 5 A 0 2 2 2 8

Leave blank

Question 6 continued

Question 6 continues on the next page

Leave blank

Question 6 continued

The assistant manager thinks that a binomial distribution is a good model but suggests that the proportion of fake coins is higher than 5%. She calculates the actual proportion of fake coins in the sample and uses this value to carry out a new hypothesis test on the data. Her expected frequencies are shown in Table 2.

Number of fake coins in each bag	0	1	2	3	4 or more
Observed frequency	43	62	26	13	6
Expected frequency	44.5	55.7	33.2	12.5	4.1

Table 2

(d) Explain why there are 2 degrees of freedom in this case.

(2)

(e) Given that she obtains a χ^2 test statistic of 2.67, test the assistant manager's hypothesis that the binomial distribution is a good model for the number of fake coins in each bag. Use a 5% level of significance and state your hypotheses clearly.

(3)

Leave blank

Question 6 continued

Q6

(Total 17 marks)

Leave blank

7. A petrol pump is tested regularly to check that the reading on its gauge is accurate. The random variable X, in litres, is the quantity of petrol actually dispensed when the gauge reads 10.00 litres. X is known to have distribution $X \sim \mathrm{N}(\mu, 0.08^2)$

(a) Eight random tests gave the following values of x

10.01 9.97 9.93 9.99 9.90 9.95 10.13 9.94

(i) Find a 95% confidence interval for μ to 2 decimal places.

(ii) Use your result to comment on the accuracy of the petrol gauge.

(5)

(b) A sample mean of 9.96 litres was obtained from a random sample of n tests. A 90% confidence interval for μ gave an upper limit of less than 10.00 litres. Find the minimum value of n.

(5)

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Q7

(Total 10 marks)

TOTAL FOR PAPER: 75 MARKS

END

P 4 3 1 5 5 A 0 2 8 2 8