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

Paper Reference(s)

6691/01

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	
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 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.
P43156A
W850/R6691/57570 5/5/1/1/

P 4 3 1 5 6 A 0 1 2 8

Turn over

PEARSON

Leave blank

1. (a) Explain what you understand by a random sample from a finite population.

(1)

(b) Give an example of a situation when it is not possible to take a random sample.

(1)

A college lecturer specialising in shoe design wants to change the way in which she organises practical work.

She decides to gather ideas from her 75 students.

She plans to give a questionnaire to a random sample of 8 of these students.

(c) (i) Describe the sampling frame that she should use.

(ii) Explain in detail how she should use a table of random numbers to obtain her sample.

(3)

Leave blank

2. The weights of pears in an orchard are assumed to have unknown mean μ and unknown standard deviation σ.

A random sample of 20 pears is taken and their weights recorded.

The sample is represented by $X_1, X_2, ..., X_{20}$. State whether or not the following are statistics. Give reasons for your answers.

(a) (i) $\dfrac{X_1 + 3X_{20}}{2}$

(ii) $\sum_{i=1}^{20} (X_i - \mu)$

(iii) $\sum_{i=1}^{20} \left(\dfrac{X_i - \mu}{\sigma} \right)$

(4)

(b) Find the mean and variance of $\dfrac{3X_1 - X_{20}}{2}$

(3)

Leave blank

3. A number of males and females were asked to rate their happiness under the headings "not happy", "fairly happy" and "very happy".

The results are shown in the table below

		Happiness			Total
		Not happy	Fairly happy	Very happy	
Gender	Female	9	43	34	86
	Male	13	25	16	54
Total		22	68	50	140

Stating your hypotheses, test at the 5% level of significance, whether or not there is evidence of an association between happiness and gender. Show your working clearly.

(10)

Leave blank

Question 3 continued

Leave blank

4. The random variable A is defined as

$$A = B + 4C - 3D$$

where B, C and D are independent random variables with

$$B \sim \mathrm{N}(6, 2^2) \qquad C \sim \mathrm{N}(7, 3^2) \qquad D \sim \mathrm{N}(4, 1.5^2)$$

Find $\mathrm{P}(A < 45)$

(6)

Leave blank

5. A research station is doing some work on the germination of a new variety of genetically modified wheat.

They planted 120 rows containing 7 seeds in each row.

The number of seeds germinating in each row was recorded. The results are as follows

Number of seeds germinating in each row	0	1	2	3	4	5	6	7
Observed number of rows	2	6	11	19	25	32	16	9

(a) Write down two reasons why a binomial distribution may be a suitable model.

(2)

(b) Show that the probability of a randomly selected seed from this sample germinating is 0.6

(2)

The research station used a binomial distribution with probability 0.6 of a seed germinating. The expected frequencies were calculated to 2 decimal places. The results are as follows

Number of seeds germinating in each row	0	1	2	3	4	5	6	7
Expected number of rows	0.20	2.06	s	23.22	t	31.35	15.68	3.36

(c) Find the value of s and the value of t.

(2)

(d) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the data can be modelled by a binomial distribution.

(7)

Leave blank

Question 5 continued

Leave blank

6. A random sample $X_1, X_2, ..., X_n$ is taken from a population with mean μ.

(a) Show that $\bar{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$ is an unbiased estimator of the population mean μ.
(1)

A company produces small jars of coffee.

Five jars of coffee were taken at random and weighed.

The weights, in grams, were as follows

197 203 205 201 195

(b) Calculate unbiased estimates of the population mean and variance of the weights of the jars produced by the company.
(3)

It is known from previous results that the weights are normally distributed with standard deviation 4.8 g.

The manager is going to take a second random sample. He wishes to ensure that there is at least a 95% probability that the estimate of the population mean is within 1.25 g of its true value.

(c) Find the minimum sample size required.
(4)

Leave blank

Question 6 continued

Leave blank

7. A machine fills packets with X grams of powder where X is normally distributed with mean μ. Each packet is supposed to contain 1 kg of powder.

To comply with regulations, the weight of powder in a randomly selected packet should be such that $P(X < \mu - 30) = 0.0005$

(a) Show that this requires the standard deviation to be 9.117 g to 3 decimal places.

(3)

A random sample of 10 packets is selected from the machine. The weight, in grams, of powder in each packet is as follows

999.8 991.6 1000.3 1006.1 1008.2 997.0 993.2 1000.0 997.1 1002.1

(b) Assuming that the standard deviation of the population is 9.117 g, test, at the 1% significance level, whether or not the machine is delivering packets with mean weight of less than 1 kg. State your hypotheses clearly.

(7)

Leave blank

Question 7 continued

Leave blank

8. The heights, in metres, and weights, in kilograms, of a random sample of 9 men are shown in the table below

Man	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Height (x)	1.68	1.74	1.75	1.76	1.78	1.82	1.84	1.88	1.98
Weight (y)	75	76	100	77	90	95	110	96	120

(a) Given that $S_{xx} = 0.0632$, $S_{yy} = 1957.5556$ and $S_{xy} = 9.3433$ calculate, to 3 decimal places, the product moment correlation coefficient between height and weight for these men.

(2)

(b) Use your value of the product moment correlation coefficient to test whether or not there is evidence of a positive correlation between the height and weight of men. Use a 5% significance level. State your hypotheses clearly.

(4)

Peter does not know the heights or weights of the 9 men. He is given photographs of them and asked to put them in order of increasing weight. He puts them in the order

A C E B G D I F H

(c) Find, to 3 decimal places, Spearman's rank correlation coefficient between Peter's order and the actual order.

(6)

(d) Use your value of Spearman's rank correlation coefficient to test for evidence of Peter's ability to correctly order men, by their weight, from their photographs. Use a 5% significance level and state your hypotheses clearly.

(4)

Leave blank

Question 8 continued

Q8

(Total 16 marks)

TOTAL FOR PAPER: 75 MARKS

END