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

Tuesday 16 June 2015 – 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	**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 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.
P44848A
W850/R6686/57570 5/1/1/1/1/1/

P 4 4 8 4 8 A 0 1 2 4

Turn over

PEARSON

1. The Sales Manager of a large chain of convenience stores is studying the sale of lottery tickets in her stores. She randomly selects 8 of her stores. From these stores she collects data for the total sales of lottery tickets in the previous January and July. The data are shown below

Store	A	B	C	D	E	F	G	H
January ticket sales (£)	1080	1639	710	1108	915	1066	1322	819
July ticket sales (£)	1113	1702	831	1048	861	1090	1303	852

(a) Use a paired t-test to determine whether or not there is evidence, at the 5% level of significance, that the mean sales of lottery tickets in this chain's stores are higher in July than in January. You should state your hypotheses and show your working clearly.

(8)

(b) State what assumption the Sales Manager needs to make about the sales of lottery tickets in her stores for the test in part (a) to be valid.

(1)

Question 1 continued

Q1

(Total 9 marks)

2. Fred is a new employee in a delicatessen. He is asked to cut cheese into 100 g blocks. A random sample of 8 of these blocks of cheese is selected. The weight, in grams, of each block of cheese is given below

94, 106, 115, 98, 111, 104, 113, 102

(a) Calculate a 90% confidence interval for the standard deviation of the weights of the blocks of cheese cut by Fred.

(6)

Given that the weights of the blocks of cheese are independent,

(b) state what further assumption is necessary for this confidence interval to be valid.

(1)

The delicatessen manager expects the standard deviation of the weights of the blocks of cheese cut by an employee to be less than 5 g. Any employee who does not achieve this target is given training.

(c) Use your answer from part (a) to comment on Fred's results.

(1)

A second employee, Olga, has just been given training. Olga is asked to cut cheese into 100 g blocks. A random sample of 20 of these blocks of cheese is selected. The weight of each block of cheese, x grams, is recorded and the results are summarised below.

$$\bar{x} = 102.6 \qquad s^2 = 19.4$$

Given that the assumption in part (b) is also valid in this case,

(d) test, at a 10% level of significance, whether or not the mean weight of the blocks of cheese cut by Olga after her training is 100 g. State your hypotheses clearly.

(6)

Question 2 continued

Question 2 continued

Question 2 continued

Q2

(Total 14 marks)

Turn over

3. As part of their research two sports science students, Ali and Bea, select a random sample of 10 adult male swimmers and a random sample of 13 adult male athletes from local sports clubs. They measure the arm span, x cm, of each person selected.
The data are summarised in the table below

	n	s^2	$\bar{x}$
Swimmers	10	48	195
Athletes	13	161	186

The students know that the arm spans of adult male swimmers and of adult male athletes may each be assumed to be normally distributed.
They decide to share out the data analysis, with Ali investigating the means of the two distributions and Bea investigating the variances of the two distributions.

Ali assumes that the variances of the two distributions are equal. She calculates the pooled estimate of variance, s_p^2

(a) Show that $s_p^2 = 112.6$ to 1 decimal place.
(2)

Ali claims that there is no difference in the mean arm spans of adult male swimmers and of adult male athletes.

(b) Stating your hypotheses clearly, test this claim at the 10% level of significance.
(5)

Bea believes that the variances of the arm spans of adult male swimmers and adult male athletes are not equal.

(c) Show that, at the 10% level of significance, the data support Bea's belief. State your hypotheses and show your working clearly.
(5)

Ali and Bea combine their work and present their results to their tutor, Clive.

(d) Explain why Clive is not happy with their research and state, with a reason, which of the tests in parts (b) and (c) is not valid.
(2)

Question 3 continued

Question 3 continued

Question 3 continued

(Total 14 marks)

Q3

4. A poultry farm produces eggs which are sold in boxes of 6. The farmer believes that the proportion, p, of eggs that are cracked when they are packed in the boxes is approximately 5%. She decides to test the hypotheses

$$H_0 : p = 0.05 \quad \text{against} \quad H_1 : p > 0.05$$

To test these hypotheses she randomly selects a box of eggs and rejects H_0 if the box contains 2 or more eggs that are cracked. If the box contains 1 egg that is cracked, she randomly selects a second box of eggs and rejects H_0 if it contains at least 1 egg that is cracked. If the first or the second box contains no cracked eggs, H_0 is immediately accepted and no further boxes are sampled.

(a) Show that the power function of this test is

$$1 - (1 - p)^6 - 6p(1 - p)^{11}$$

(3)

(b) Calculate the size of this test.

(2)

Given that $p = 0.1$

(c) find the expected number of eggs inspected each time this test is carried out, giving your answer correct to 3 significant figures,

(3)

(d) calculate the probability of a Type II error.

(2)

Given that $p = 0.1$ is an unacceptably high value for the farmer,

(e) use your answer from part (d) to comment on the farmer's test.

(1)

Question 4 continued

Question 4 continued

Question 4 continued

Q4

(Total 11 marks)

5. A researcher is investigating the accuracy of IQ tests. One company offers IQ tests that it claims will give any individual's IQ with a standard deviation of 5

The researcher takes these tests 9 times with the following results

123, 118, 127, 120, 134, 120, 118, 135, 121

(a) Find the sample mean, $\overline{x}$, and the sample variance, s^2, of these scores.

(2)

Given that any individual's IQ scores on these tests are independent and have a normal distribution,

(b) use the hypotheses

$$H_0 : \sigma^2 = 25 \quad \text{against} \quad H_1 : \sigma^2 > 25$$

to test the company's claim at the 5% significance level.

(4)

Gurdip works for the company and has taken these IQ tests 12 times. Gurdip claims that the sample variance of these 12 scores is $s^2 = 8.17$

(c) Use this value of s^2 to calculate a 95% confidence interval for the variance of Gurdip's IQ test scores.

[You may use $P(\chi^2_{11} > 3.816) = 0.975$ and $P(\chi^2_{11} > 21.920) = 0.025$]

(2)

(d) Assuming that $\sigma^2 = 25$, comment on Gurdip's claim.

(1)

Question 5 continued

Question 5 continued

Question 5 continued

(Total 9 marks)

Q5

6. A random sample $X_1, X_2, X_3, \ldots, X_{2n}$ is taken from a population with mean $\frac{\mu}{3}$ and variance $3\sigma^2$. A second random sample $Y_1, Y_2, Y_3, \ldots, Y_n$ is taken from a population with mean $\frac{\mu}{2}$ and variance $\frac{\sigma^2}{2}$, where the X and Y variables are all independent.

A, B and C are possible estimators of μ, where

$$A = \frac{X_1 + X_2 + X_3 + Y_1 + Y_2}{2}$$

$$B = \frac{3X_1}{2} + \frac{2Y_1}{3}$$

$$C = \frac{3X_1 + 4Y_1}{3}$$

(a) Show that two of A, B and C are unbiased estimators of μ and find the bias of the third estimator of μ.
(5)

(b) Showing your working clearly, find which of A, B and C is the best estimator of μ.
(4)

The estimator

$$D = \frac{1}{k}\left(\sum_{i=1}^{2n} X_i + \sum_{i=1}^{n} Y_i\right)$$

is an unbiased estimator of μ.

(c) Find k in terms of n.
(3)

(d) Show that D is also a consistent estimator of μ.
(4)

(e) Find the least value of n for which D is a better estimator of μ than any of A, B or C.
(2)

Question 6 continued

Question 6 continued

Question 6 continued

Question 6 continued

Q6

(Total 18 marks)

TOTAL FOR PAPER: 75 MARKS

END