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

Thursday 24 June 2010 – 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	
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, differentiation and 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.
M35398A
W850/R6686/57570 4/5/5/3
M 3 5 3 9 8 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

1. A teacher wishes to test whether playing background music enables students to complete a task more quickly. The same task was completed by 15 students, divided at random into two groups. The first group had background music playing during the task and the second group had no background music playing.
The times taken, in minutes, to complete the task are summarised below.

	Sample size n	Standard deviation s	Mean $\bar{x}$
With background music	8	4.1	15.9
Without background music	7	5.2	17.9

You may assume that the times taken to complete the task by the students are two independent random samples from normal distributions.

(a) Stating your hypotheses clearly, test, at the 10% level of significance, whether or not the variances of the times taken to complete the task with and without background music are equal.

(5)

(b) Find a 99% confidence interval for the difference in the mean times taken to complete the task with and without background music.

(7)

Experiments like this are often performed using the same people in each group.

(c) Explain why this would not be appropriate in this case.

(1)

Question 1 continued

2. As part of an investigation, a random sample of 10 people had their heart rate, in beats per minute, measured whilst standing up and whilst lying down. The results are summarised below.

Person	1	2	3	4	5	6	7	8	9	10
Heart rate lying down	66	70	59	65	72	66	62	69	56	68
Heart rate standing up	75	76	63	67	80	75	65	74	63	75

(a) State one assumption that needs to be made in order to carry out a paired t-test.

(1)

(b) Test, at the 5% level of significance, whether or not there is any evidence that standing up increases people's mean heart rate by more than 5 beats per minute. State your hypotheses clearly.

(8)

Question 2 continued

3. A manager in a sweet factory believes that the machines are working incorrectly and the proportion p of underweight bags of sweets is more than 5%. He decides to test this by randomly selecting a sample of 5 bags and recording the number X that are underweight. The manager sets up the hypotheses H_0: $p = 0.05$ and H_1: $p > 0.05$ and rejects the null hypothesis if $x > 1$.

(a) Find the size of the test.

(2)

(b) Show that the power function of the test is

$$1 - (1 - p)^4(1 + 4p)$$

(3)

The manager goes on holiday and his deputy checks the production by randomly selecting a sample of 10 bags of sweets. He rejects the hypothesis that $p = 0.05$ if more than 2 underweight bags are found in the sample.

(c) Find the probability of a Type I error using the deputy's test.

(2)

Question 3 continues on page 12

Question 3 continued

Question 3 continues on the next page

Question 3 continued

The table below gives some values, to 2 decimal places, of the power function for the deputy's test.

p	0.10	0.15	0.20	0.25
Power	0.07	s	0.32	0.47

(d) Find the value of s.

(1)

The graph of the power function for the manager's test is shown in Figure 1.

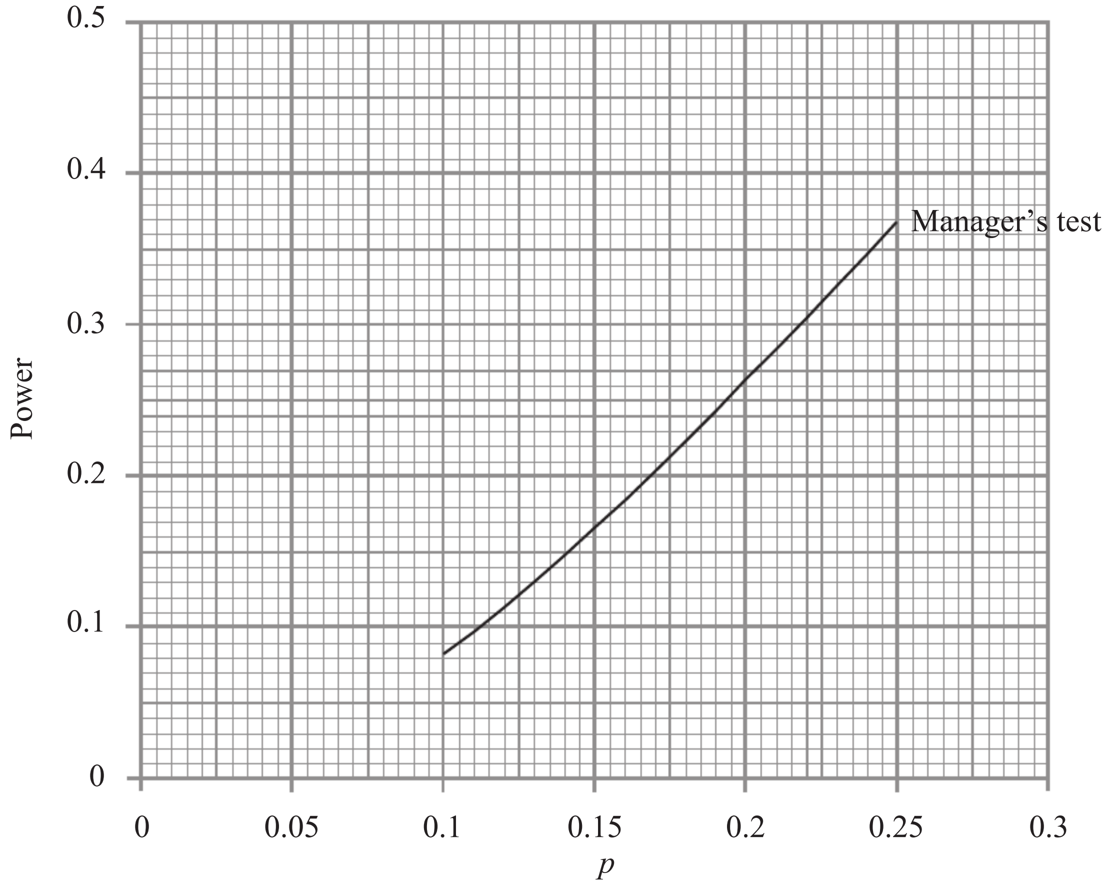

Figure 1

(e) On the same axes, draw the graph of the power function for the deputy's test.

(1)

(f) (i) State the value of p where these graphs intersect.

(ii) Compare the effectiveness of the two tests if p is greater than this value.

(2)

The deputy suggests that they should use his sampling method rather than the manager's.

(g) Give a reason why the manager might not agree to this change.

(1)

Question 3 continued

Q3

(Total 12 marks)

4. A random sample of 15 strawberries is taken from a large field and the weight x grams of each strawberry is recorded. The results are summarised below.

$$\sum x = 291 \qquad \sum x^2 = 5968$$

Assume that the weights of strawberries are normally distributed.
Calculate a 95% confidence interval for

(a) (i) the mean of the weights of the strawberries in the field,

(ii) the variance of the weights of the strawberries in the field.

(12)

Strawberries weighing more than 23g are considered to be less tasty.

(b) Use appropriate confidence limits from part (a) to find the highest estimate of the proportion of strawberries that are considered to be less tasty.

(4)

Question 4 continued

5. A car manufacturer claims that, on a motorway, the mean number of miles per gallon for the Panther car is more than 70. To test this claim a car magazine measures the number of miles per gallon, x, of each of a random sample of 20 Panther cars and obtained the following statistics.

$$\bar{x} = 71.2 \qquad s = 3.4$$

The number of miles per gallon may be assumed to be normally distributed.

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

The standard deviation of the number of miles per gallon for the Tiger car is 4.

(b) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is evidence that the variance of the number of miles per gallon for the Panther car is different from that of the Tiger car.
(6)

Question 5 continued

6. Faults occur in a roll of material at a rate of λ per m^2. To estimate λ, three pieces of material of sizes $3\,m^2$, $7\,m^2$ and $10\,m^2$ are selected and the number of faults X_1, X_2 and X_3 respectively are recorded.

The estimator $\hat{\lambda}$, where

$$\hat{\lambda} = k(X_1 + X_2 + X_3)$$

is an unbiased estimator of λ.

(a) Write down the distributions of X_1, X_2 and X_3 and find the value of k.

(4)

(b) Find $\text{Var}(\hat{\lambda})$.

(3)

A random sample of n pieces of this material, each of size $4\,m^2$, was taken. The number of faults on each piece, Y, was recorded.

(c) Show that $\frac{1}{4}\bar{Y}$ is an unbiased estimator of λ.

(2)

(d) Find $\text{Var}(\frac{1}{4}\bar{Y})$.

(3)

(e) Find the minimum value of n for which $\frac{1}{4}\bar{Y}$ becomes a better estimator of λ than $\hat{\lambda}$.

(2)

Question 6 continued