Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
TOTAL	

General Certificate of Education
Advanced Level Examination
June 2013

Mathematics MS04

Unit Statistics 4

Friday 21 June 2013 9.00 am to 10.30 am

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer each question in the space provided for that question. If you require extra space, use an AQA supplementary answer book; do **not** use the space provided for a different question.
- Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.
- You do not necessarily need to use all the space provided.

Do not write outside the box

Answer **all** questions.

Answer each question in the space provided for that question.

1 A gardener decided to compare the variability in the heights of sunflowers grown on the sunny side of her garden with the variability in the heights of those grown on the shady side of her garden.

She selected a sample of 13 sunflowers on the sunny side and a sample of 10 sunflowers on the shady side of her garden. She measured the height of each sunflower, in metres, correct to two decimal places. Each sample may be regarded as a random sample.

The independent random variables X and Y denote the height, in metres, of sunflowers grown on the sunny side and on the shady side of her garden respectively. You may assume that X and Y are normally distributed with variances σ_X^2 and σ_Y^2 respectively.

The following results were obtained.

$$\sum(x-\bar{x})^2 = 4.68 \quad \text{and} \quad \sum(y-\bar{y})^2 = 8.10$$

(a) Calculate unbiased estimates of σ_X^2 and σ_Y^2. *(1 mark)*

(b) The gardener believed that the height of sunflowers grown on the shady side was more variable than the height of sunflowers grown on the sunny side.

Carry out an F-test, using a significance level of 5%, to test this belief. *(6 marks)*

QUESTION PART REFERENCE

Answer space for question 1

Do not write outside the box

QUESTION PART REFERENCE	Answer space for question 1

2 The random variable X has a geometric distribution with parameter p. It is given that $\mathrm{Var}(X)$ is four times $\mathrm{E}(X)$.

(a) Show that the non-zero value of p is 0.20. *(3 marks)*

(b) Hence find:

(i) $\mathrm{P}(X > 7 \mid X > 4)$; *(3 marks)*

(ii) the least integer n such that $\mathrm{P}(X > n) < 0.0001$. *(4 marks)*

QUESTION PART REFERENCE

Answer space for question 2

QUESTION PART REFERENCE

Answer space for question 2

3 The differences, d_i, between 8 pairs of readings in a paired-sample t-test gave $\overline{d} = 0.1625$. The result of this test, which was one-tailed, was just significant at the 10% level of significance. The conclusion was therefore that one mean value was likely to be greater than the other.

(a) Find the value of s that was used in the test, where s^2 is an unbiased estimate of σ^2, the variance of the population of differences. *(5 marks)*

(b) Hence construct a 95% confidence interval for σ. *(5 marks)*

QUESTION PART REFERENCE

Answer space for question 3

Do not write outside the box

QUESTION PART REFERENCE

Answer space for question 3

Do not write outside the box

4 The random variable X has an exponential distribution with mean μ. The cumulative distribution function of X for $x \geqslant 0$ is given by

$$F(x) = 1 - e^{-\frac{1}{\mu}x}$$

(a) Find an exact expression for the interquartile range of X in terms of μ. *(5 marks)*

(b) Prove, by integration, that $E(X^2) = 2\mu^2$. *(4 marks)*

(c) Show that the standard deviation of X is less than the interquartile range of X. *(3 marks)*

QUESTION PART REFERENCE

Answer space for question 4

QUESTION PART REFERENCE

Answer space for question 4

QUESTION PART REFERENCE

Answer space for question 4

Do not write outside the box

QUESTION PART REFERENCE

Answer space for question 4

Turn over ▶

Do not write outside the box

5 An office worker believed that, on average, his journey time from home to the office in the morning was longer than his journey time back home in the evening. In order to test this belief, he recorded his journey times, to the nearest minute, for a random sample of 8 morning journeys and a random sample of 10 evening journeys. The results are summarised below.

Mornings	$n = 8$	$\sum m = 204$	$\sum(m - \overline{m})^2 = 470$
Evenings	$n = 10$	$\sum e = 180$	$\sum(e - \overline{e})^2 = 300$

(a) State **two** assumptions that need to be made so that an independent-samples t-test is valid. *(2 marks)*

(b) Making these assumptions, investigate, at the 2.5% level of significance, the office worker's belief. *(9 marks)*

QUESTION PART REFERENCE

Answer space for question 5

QUESTION PART REFERENCE

Answer space for question 5

6 Gwyneth, a retired maths teacher, believes that the number of emails that she receives each day can be modelled by a Poisson distribution. She records the number of emails received each day for a random sample of 100 days. The results that she obtains are as follows.

Number of emails per day	0	1	2	3	4	5	6
Frequency	11	28	25	16	9	8	3

(a) Calculate the mean number of emails that Gwyneth received each day. *(1 mark)*

(b) Complete the following table of expected frequencies for a Poisson distribution with mean equal to that calculated in part **(a)**.

Number of emails per day	0	1	2	3	4	5	$\geqslant 6$
Expected frequency	11.08	24.38	26.81				

(3 marks)

(c) Perform a χ^2-test, at the 10% level of significance, to investigate Gwyneth's belief. *(7 marks)*

QUESTION PART REFERENCE

Answer space for question 6

QUESTION PART REFERENCE

Answer space for question 6

Do not write outside the box

QUESTION PART REFERENCE

Answer space for question 6

Do not write outside the box

QUESTION PART REFERENCE	Answer space for question 6

Turn over ▶

Do not write outside the box

7 The conductivity, γ, of metal wire is estimated by observing a related random variable, R, which has probability density function

$$f(r) = \begin{cases} \dfrac{2r}{\gamma^2} & 0 \leqslant r \leqslant \gamma \\ 0 & \text{otherwise} \end{cases}$$

(a) (i) Show that $\frac{3}{2}R$ is an unbiased estimator of γ. *(4 marks)*

(ii) Given that the variance of R is $\frac{1}{18}\gamma^2$, find, in terms of γ, the variance of $\frac{3}{2}R$. *(2 marks)*

(b) The conductivity can also be estimated by making observations of a random variable, S, which has mean $\frac{1}{4}\gamma$ and variance $\frac{1}{16}\gamma^2$. The random variable T is defined by $T = S_1 + S_2 + S_3$, where S_1, S_2 and S_3 are three independent observations of S.

(i) Find the value of the constant k such that kT is an unbiased estimator of γ. *(2 marks)*

(ii) Hence find the relative efficiency of $\frac{3}{2}R$ with respect to kT. *(4 marks)*

(iii) State, with justification, which of $\frac{3}{2}R$ and kT is a preferred unbiased estimator of γ. *(2 marks)*

QUESTION PART REFERENCE

Answer space for question 7

QUESTION PART REFERENCE

Answer space for question 7

Turn over ▶

Do not write outside the box

QUESTION PART REFERENCE

Answer space for question 7

END OF QUESTIONS