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 2011

Mathematics MS04

Unit Statistics 4

Thursday 23 June 2011 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 the questions in the spaces provided. 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.

Answer **all** questions in the spaces provided.

1 A random sample of 12 batches of steel was taken from a production line. The carbon content, measured in grams per cubic metre, of each of these batches of steel is given below.

3.6 4.3 3.8 4.1 3.7 3.6 3.9 4.2 4.3 4.1 4.4 3.5

Assuming that this sample came from an underlying normal population, test, at the 1% level of significance, whether the standard deviation of the carbon content of steel from this production line is 0.7. *(7 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Do not write outside the box

2 The fat content of minced lamb, measured in grams, can be determined by two different procedures, A and B. A random sample of 8 equal portions of minced lamb is selected and each portion is divided into two halves. Procedure A is applied to one half and Procedure B is applied to the other half.

The results are shown in the table.

	Fat content of minced lamb (grams)							
Procedure A	24	70	48	18	36	45	62	45
Procedure B	22	65	47	18	32	51	59	41

(a) Given that

$D =$ fat content as determined by Procedure A $-$ fat content as determined by Procedure B

use the above data to determine a 99% confidence interval for the mean value of D. *(8 marks)*

(b) State an assumption that you have made about the distribution of D. *(1 mark)*

(c) Comment on a suggestion that the mean value of D is 5. *(2 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

3 The length of time, X days, for rust to appear on cars treated with *Killrust* may be assumed to be normally distributed with variance $\sigma_X{}^2$. The length of time, Y days, for rust to appear on cars treated with *Stoprust* may be assumed to be normally distributed with variance $\sigma_Y{}^2$.

The lengths of time, x days, measured to the nearest ten days, for rust to appear on a random sample of 10 cars treated with *Killrust* were

2500 2690 2390 2680 2800 2700 2470 2580 2610 2650

The lengths of time, y days, measured to the nearest ten days, for rust to appear on a random sample of 8 cars treated with *Stoprust* were

2620 2500 2520 2420 2460 2490 2590 2580

(a) Calculate unbiased estimates of $\sigma_X{}^2$ and $\sigma_Y{}^2$. *(2 marks)*

(b) (i) Determine a 98% confidence interval for the variance ratio $\dfrac{\sigma_X{}^2}{\sigma_Y{}^2}$. *(7 marks)*

(ii) Hence comment on the suggestion that the rust-free period of time for cars treated with *Killrust* is more variable than that for cars treated with *Stoprust*. *(2 marks)*

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

4 A cosmologist claimed that the lifetime of a certain particle, measured in picoseconds, can be modelled by the random variable T, which has cumulative distribution function

$$\mathrm{F}(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2}t^3 - \frac{3}{16}t^4 & 0 \leqslant t \leqslant 2 \\ 1 & t > 2 \end{cases}$$

To test this claim, the cosmologist first divided the interval $[0, 2]$ into five equal intervals and then recorded into which of these intervals the lifetimes of 50 randomly selected such particles fell.

The results are shown in the table.

Interval	**0 – 0.4**	**0.4 – 0.8**	**0.8 – 1.2**	**1.2 – 1.6**	**1.6 – 2.0**
Number of particles	2	9	12	22	5

(a) Assuming that the cosmologist's claim is correct:

(i) evaluate $\mathrm{F}(t)$ for $t = 0.4$, 0.8, 1.2, 1.6 and 2.0;

(ii) complete the table opposite. *(4 marks)*

(b) Hence use a χ^2 test, at the 5% level of significance, to investigate the cosmologist's claim. *(9 marks)*

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

(a)(ii)

Interval	0 – 0.4	0.4 – 0.8	0.8 – 1.2	1.2 – 1.6	1.6 – 2.0
Probability	0.0272	0.1520			

Turn over ▶

Do not write outside the box

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

5 (a) The random variable X follows a geometric distribution.

Show that $\mathrm{E}(X) = \dfrac{1}{p}$, where p is the probability of success in a single trial.

(3 marks)

(b) Andy plays a game with Bea by throwing an unbiased six-sided die until a 5 is obtained. When a 5 is obtained, Bea pays Andy £10.

Find, to the nearest penny, the amount that Bea should charge Andy **per throw** so that, in the long run, Bea makes a profit of £1 **per game**. *(4 marks)*

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

6 (a) The continuous random variable X follows an exponential distribution if it has a probability density function

$$f(x) = \begin{cases} ke^{-kx} & x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

where k is a positive constant.

(i) Prove that the mean value, μ, of X is $\dfrac{1}{k}$. *(3 marks)*

(ii) Find, in terms of k, the median value, m, of X and hence show that $m < \mu$. *(5 marks)*

(b) The number of radioactive particles striking a screen in a time period of length t seconds follows a Poisson distribution with mean $\dfrac{t}{\lambda}$, where λ is a constant.

(i) Write down the probability that no particles strike the screen in a period of t seconds. *(1 mark)*

(ii) The random variable T is defined as the length of time, in seconds, between successive radioactive particles striking the screen.

(A) Show that

$$P(T < t) = 1 - e^{-\frac{t}{\lambda}}$$ *(2 marks)*

(B) Hence, by finding the probability density function of T, state the distribution of T. *(2 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Do not write outside the box

7 (a) The statistic T is derived from a random sample taken from a population which has an unknown parameter θ. T is an unbiased estimator for θ.

What does the statement "T is an unbiased estimator for θ" imply? *(1 mark)*

(b) A random sample of size n is taken from each of two independent populations.

The first population has mean μ and variance σ^2, and $\overline{X}$ denotes the sample mean.

The second population has mean $\dfrac{\mu}{3}$ and variance $b\sigma^2$, where b is a positive constant, and $\overline{Y}$ denotes the sample mean.

Two unbiased estimators for μ are defined by

$$T_1 = 4\overline{X} - a\overline{Y} \qquad \text{and} \qquad T_2 = \tfrac{1}{9}(8\overline{X} + 3\overline{Y})$$

(i) Determine the value of a. *(3 marks)*

(ii) Show that $\mathrm{Var}(T_1) = \dfrac{\sigma^2}{n}(16 + 81b)$ and find a simplified expression for $\mathrm{Var}(T_2)$. *(5 marks)*

(iii) Calculate the relative efficiency of T_2 with respect to T_1 and decide, giving a reason, which of T_1 or T_2 is the more efficient estimator for μ. *(4 marks)*

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

END OF QUESTIONS