Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

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

General Certificate of Education
Advanced Level Examination
June 2012

Mathematics MS04

Unit Statistics 4

Monday 25 June 2012 1.30 pm to 3.00 pm

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.

Answer **all** questions.

Answer each question in the space provided for that question.

1 An angler is investigating the breaking strength of nylon fishing line. The angler collects samples from each of two retail outlets: *Hot Rods* and *The Reel Deal*. For each sample, he measures the breaking strengths in kilograms. The results are shown in the table.

Hot Rods	6.314	6.242	6.725	6.782	6.582	6.173	6.467	5.830	6.145
The Reel Deal	6.435	6.403	6.155	7.074	6.709	6.540	6.303	6.768	

Assuming that the measurements are independent random samples from two normal distributions having the same variance, test, at the 10% level of significance, the assertion that the mean breaking strength of nylon fishing line is the same for line supplied by *Hot Rods* as it is for line supplied by *The Reel Deal*. *(10 marks)*

QUESTION PART REFERENCE

Answer space for question 1

QUESTION PART REFERENCE	Answer space for question 1

Turn over ▶

2 A psychologist recorded the reaction times, in seconds, of Subject A, who was exposed to identical stimuli on 10 random occasions. The times were

0.40 0.42 0.39 0.40 0.41 0.42 0.43 0.44 0.43 0.42

(a) Assuming that these times came from an underlying normal population, construct a 98% confidence interval for the population standard deviation. *(6 marks)*

(b) The psychologist also recorded the reaction times, in seconds, of Subject B, to the same stimuli under the same conditions as for Subject A, with the following results:

0.42 0.49 0.46 0.42 0.48 0.44 0.49 0.44

Assuming that these times also came from an underlying normal population, test the hypothesis, at the 10% level of significance, that the reaction times of Subject A and those of Subject B have the same variance. *(7 marks)*

QUESTION PART REFERENCE

Answer space for question 2

QUESTION PART REFERENCE	Answer space for question 2

3 A horticulturalist was considering the germination rate of a large batch of runner-bean seeds. He selected a random sample of 500 seeds and planted them in 100 rows of 5 seeds. He then recorded the number of seeds germinating in each of the 100 rows. The results that he obtained are shown in the table.

Number of seeds germinating (x)	0	1	2	3	4	5
Number of rows (f)	25	41	20	12	2	0

(a) Calculate the mean number of seeds germinating per row, and hence show that an estimated value for p, the probability that a seed germinates, is 0.25 . *(2 marks)*

(b) The model $\mathrm{B}(5, 0.25)$ is suggested as suitable for the number of seeds germinating per row. Calculate the expected frequencies for this model. *(4 marks)*

(c) Use a χ^2 goodness of fit test, at the 5% level of significance, to investigate whether the model $\mathrm{B}(5,\ p)$ is suitable for the number of seeds germinating per row. *(7 marks)*

(d) Does your conclusion in part **(c)** support the view that the probability that a seed germinates is the same whichever row it is planted in? Explain your answer. *(2 marks)*

QUESTION PART REFERENCE

Answer space for question 3

QUESTION PART REFERENCE

Answer space for question 3

QUESTION PART REFERENCE

Answer space for question 3

QUESTION PART REFERENCE

Answer space for question 3

4 (a) State why, in choosing between two unbiased estimators of a parameter, the one with the smaller variance is preferred. *(2 marks)*

(b) The random variable $\overline{X}_1$ denotes the mean of a random sample of size n_1 taken from a normal population with mean μ_1 and variance $\sigma_1{}^2$. The random variable $\overline{X}_2$ denotes the mean of a random sample of size n_2 taken from an independent normal population with mean μ_2 and variance $\sigma_2{}^2$.

(i) Show that $\overline{X}_1 - \overline{X}_2$ is an unbiased estimator of $\mu_1 - \mu_2$ and write down the variance of this estimator. *(3 marks)*

(ii) It is given that $n_1 + n_2 = n$, where n is a fixed number, and that n_1 and n_2 are so large that they may be assumed to be continuous variables. Given further that the variance of $\overline{X}_1 - \overline{X}_2$ has a minimum value, show that this minimum value occurs when $n_1 : n_2 = \sigma_1 : \sigma_2$. *(4 marks)*

(iii) Find the values of n_1 and n_2 which minimise the variance of $\overline{X}_1 - \overline{X}_2$ in the case when $\sigma_1{}^2 = 0.0025$, $\sigma_2{}^2 = 0.0081$ and $n = 280$. *(2 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

QUESTION PART REFERENCE	Answer space for question 4

Turn over ▶

5 A random variable X has an exponential distribution with probability density function $f(x)$, where

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

and k is a constant.

(a) Given that $E(X) = \frac{1}{k}$, find:

(i) using integration, $E(X^2)$;

(ii) $\text{Var}(X)$. *(6 marks)*

(b) (i) Derive the cumulative distribution function, $F(x)$, of X for $x \geqslant 0$. *(3 marks)*

(ii) Hence find, in terms of k, the **exact** value of the 90th percentile of X. *(3 marks)*

(c) A machine has two essential components, the lifetimes of which follow exponential distributions with means a hours and $3a$ hours. The machine will stop if either component fails. The failures of the two components may be taken to be independent.

Find the probability that the machine continues to work for at least a hours from the start, giving your answer in the form e^q, where q is a rational number to be determined. *(4 marks)*

QUESTION PART REFERENCE	**Answer space for question 5**

QUESTION PART REFERENCE	**Answer space for question 5**

QUESTION PART REFERENCE	Answer space for question 5

QUESTION PART REFERENCE	Answer space for question 5

6 (a) The random variable X has a geometric distribution with parameter p.

(i) Prove that $E(X) = \frac{1}{p}$. *(3 marks)*

(ii) Given that $E(X^2) = \frac{2-p}{p^2}$, show that $\text{Var}(X) = \frac{1-p}{p^2}$. *(2 marks)*

(b) An unbiased tetrahedral die has faces marked 1 to 4. When it is thrown on a table, the score is the number on the face that is in contact with the table.

(i) Calculate the probability that it takes more than two throws to obtain a score of 4. *(2 marks)*

(ii) The number of throws, Y, that it takes for **two different** scores to occur at least once is given by $Y = 1 + X$, where X has a geometric distribution with parameter $\frac{3}{4}$.

Determine values for $E(Y)$ and $\text{Var}(Y)$. *(3 marks)*

QUESTION PART REFERENCE	**Answer space for question 6**

QUESTION PART REFERENCE

Answer space for question 6

Turn over ▶

QUESTION PART REFERENCE

Answer space for question 6

END OF QUESTIONS