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 2010

Mathematics MS04

Unit Statistics 4

Thursday 24 June 2010 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.

Do not write outside the box

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

1 A random sample of 8 male 400-metre runners was selected and their personal best times, in seconds, achieved at sea level and at altitude were recorded, as shown in the table.

Runner	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Sea level time	48.4	47.6	48.4	45.9	47.1	50.4	48.2	50.8
Altitude time	47.9	47.1	47.7	45.7	46.8	50.3	47.9	50.3

A coach believes that the personal best times of male 400-metre runners at sea level will be more than 0.2 seconds slower than those at altitude.

Assuming that the differences between times at sea level and at altitude are normally distributed, investigate the coach's belief at the 1% level of significance. *(10 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Do not write outside the box

2 (a) A different Sudoku puzzle appears on television each day. Every day, Ethan solves the puzzle and he records the time taken, x minutes, to solve each puzzle during a 10-day period with the following results, correct to one decimal place.

24.2 27.3 22.4 28.1 26.4 24.9 30.6 23.8 27.7 24.0

He models these times as independent values of a normally distributed random variable with variance σ^2.

Determine a 95% confidence interval for σ. *(6 marks)*

(b) Nadia also solves the Sudoku puzzle which appears on television each day. She believes that her times taken to solve puzzles are less variable than the times taken by Ethan.

To test this belief, she records her time taken, y minutes, to solve each puzzle during a 10-day period. Her results, correct to one decimal place, are as follows.

16.9 15.2 18.2 14.6 17.3 16.0 18.1 14.8 16.5 14.9

You may assume that these times are independent and are normally distributed and that

$$\sum(y - \bar{y})^2 = 16.625$$

(i) State appropriate hypotheses to test Nadia's belief.

(ii) Use an F-test, with a significance level of 5%, to test Nadia's belief. *(7 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

3 A certain species of snake occurs in Africa and in Asia. It is thought that adult snakes of this species found in Africa are, on average, longer than those found in Asia.

Two small samples of adult snakes of this species were taken, one from Africa and one from Asia. Data relating to the lengths, in metres, of the snakes in each sample are summarised in the table below.

African lengths (*x* metres)	**Asian lengths (*y* metres)**
$\bar{x} = 0.6386$	$\bar{y} = 0.3200$
$\sum(x - \bar{x})^2 = 0.2958$	$\sum(y - \bar{y})^2 = 0.1873$
Sample size $= 7$	Sample size $= 6$

(a) Construct a 99% confidence interval for $\mu_X - \mu_Y$, the difference in mean length between African and Asian adult snakes of this species. *(7 marks)*

(b) State **three** assumptions that you have made in constructing this confidence interval. *(3 marks)*

(c) State, with a reason, whether African adult snakes of the species are, on average, longer than Asian adult snakes of the species. *(2 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

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

(i) Prove from first principles that $\mathrm{E}(X(X-1)) = \dfrac{2(1-p)}{p^2}$. *(3 marks)*

(ii) Hence, given that $\mathrm{E}(X) = \dfrac{1}{p}$, prove that $\mathrm{Var}(X) = \dfrac{1-p}{p^2}$. *(2 marks)*

(b) The independent random variables X_1 and X_2 have geometric distributions with parameters p_1 and p_2 respectively.

It is given that $\dfrac{\mathrm{E}(X_1)}{\mathrm{E}(X_2)} = \dfrac{2}{3}$ and that $\dfrac{\mathrm{Var}(X_1)}{\mathrm{Var}(X_2)} = \dfrac{1}{3}$.

(i) Show that $p_1 = \dfrac{1}{2}$. *(5 marks)*

(ii) Hence find the least value of N such that $\mathrm{P}(X_1 > N) < 10^{-5}$. *(4 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

QUESTION PART REFERENCE

5 The random variable X may be modelled by an exponential distribution with probability density function

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

(a) Show that, for $x \geqslant 0$, the cumulative distribution function is given by

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

(2 marks)

(b) For an exponential distribution with $\lambda = 0.5$, find values for the two probabilities missing from the following table, giving your answers to four decimal places.

Interval	**0–1**	**1–2**	**2–3**	**3–4**	**4–5**	**$\geqslant$5**
Probability	0.3935	0.2387	0.1447	0.0878		

(2 marks)

(c) An engineering company manufactures mechanical components for power stations. A random sample of 80 of these components was selected. The length of time, in years, to the failure of each component was recorded. These lifetimes are shown in the table.

Lifetime (years)	**0–1**	**1–2**	**2–3**	**3–4**	**4–5**	**$\geqslant$5**
Number of components	34	20	9	6	2	9

Use a χ^2 test to determine, at the 10% level of significance, whether the exponential distribution with $\lambda = 0.5$ is a suitable model for the lifetime, in years, of these mechanical components. *(8 marks)*

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

QUESTION PART REFERENCE

6 The random variable X has a probability density function

$$f(x) = \begin{cases} \frac{1}{4}\sin\left(\frac{x}{2}\right) & 0 \leqslant x \leqslant 2\pi \\ 0 & \text{otherwise} \end{cases}$$

For this distribution, you are given that

$$E(X) = \pi \quad \text{and} \quad E(X^2) = 2\pi^2 - 8$$

(a) Find, in terms of π, the variance of X. *(2 marks)*

(b) The mean of a random sample of n observations, $X_1, X_2, X_3, \ldots, X_n$, is denoted by $\overline{X}$.

(i) Write down, in terms of π and n, expressions for the mean and the variance of $\overline{X}$. *(2 marks)*

(ii) Explain why $\overline{X}$ is an unbiased and consistent estimator for π. *(2 marks)*

(c) (i) For a random sample of size 5, the median, M, is an unbiased estimator for π with variance $\pi^2 - \dfrac{2072}{225}$.

For such a sample, calculate the relative efficiency of M with respect to $\overline{X}$, giving your answer to three decimal places.

Hence give a reason why $\overline{X}$ should be preferred to M as an estimator for π. *(3 marks)*

(ii) A particular random sample of 5 observations of X gave the following values.

0.80 1.99 3.12 3.89 6.20

(A) Given that $P(X \geqslant 2\pi) = 0$, use this sample to find an inequality for π.

(B) Obtain the values of $\bar{x}$ and m for this sample.

(C) Comment on your answers in part **(c)(ii)(B)** in the light of your answers to part **(c)(i)**. *(5 marks)*

Do not write outside the box

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

END OF QUESTIONS