General Certificate of Education
June 2004
Advanced Level Examination

MATHEMATICS (SPECIFICATION A) **MAS4/W**
Unit Statistics 4

Tuesday 29 June 2004 Afternoon Session

In addition to this paper you will require:
- an 8-page answer book;
- one sheet of graph paper for use in Question 4;
- a ruler;
- the AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed: 1 hour 20 minutes

Instructions
- Use blue or black ink or ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The *Examining Body* for this paper is AQA. The *Paper Reference* is MAS4/W.
- Answer **all** questions.
- All necessary working should be shown; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.
- Tie loosely any additional sheets you have used to the back of your answer book before handing it to the invigilator.

Information
- The maximum mark for this paper is 60.
- Mark allocations are shown in brackets.
- Additional sheets of graph paper are available on request.

Advice
- Unless stated otherwise, formulae may be quoted, without proof, from the booklet.

Answer **all** questions.

1 The table below shows the marks awarded by two judges to eight competitors in a poetry competition.

Competitor	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Judge 1	46	42	33	57	42	38	54	42
Judge 2	56	47	35	32	51	45	40	38

(a) Calculate the value of Spearman's rank correlation coefficient between the marks of Judge 1 and the marks of Judge 2. *(5 marks)*

(b) The value of Spearman's rank correlation coefficient between a third judge, Judge 3, and Judge 2 is -1.

Give the positions of the competitors according to Judge 3. *(2 marks)*

(c) In the light of the evidence available, comment on the ease of selecting a winner. *(2 marks)*

2 A police authority conducts an eight week experiment. In each week it records the number of foot patrols, x, made in a small town and the number of reported crimes, y, in that town. The data are summarised as follows.

$$\sum x = 52 \qquad \sum x^2 = 380 \qquad \sum xy = 1335$$

$$\sum y = 225 \qquad \sum y^2 = 7007 \qquad n = 8$$

(a) Calculate the value of the product moment correlation coefficient for these data. *(5 marks)*

(b) Assuming that these data are a random sample from a distribution with correlation coefficient ρ, investigate, at the 2.5% level of significance, the hypothesis that $\rho < 0$. *(4 marks)*

(c) In view of your conclusion in part (b), what advice should be given to the Chief Constable of this police authority? *(1 mark)*

3 The probability that a particular medicine cures a certain disease in any individual is θ.

The medicine is given to n individuals with the disease and a proportion p are cured.

(a) State, in terms of θ and n, the mean and variance of the probability distribution for p. *(2 marks)*

(b) State **two** conditions which would allow this probability distribution to be approximated by a normal distribution. *(2 marks)*

(c) The medicine is given to 200 individuals with the disease and 180 of these individuals are cured.

Construct an approximate 95% confidence interval for θ. *(4 marks)*

4 [A sheet of graph paper is provided for use in answering this question.]

A mathematics teacher recorded the length of time, y minutes, taken to travel to school when leaving home x minutes after 7 am on seven selected mornings. The results are as follows.

x	0	10	20	30	40	50	60
y	16	27	28	39	39	48	51

(a) Plot the data on a scatter diagram. *(3 marks)*

(b) (i) Calculate the equation of the least squares regression line of y on x, writing your answer in the form $y = a + bx$. *(5 marks)*

(ii) Draw the regression line on your scatter diagram. *(1 mark)*

(c) The mathematics teacher needs to arrive at school no later than 8.40 am.

The number of minutes by which the mathematics teacher arrives early at school, when leaving home x minutes after 7 am, is denoted by z.

(i) Deduce that

$$z = (100 - a) - (1 + b)x.$$

(3 marks)

(ii) Hence estimate, to the nearest minute, the latest time that the mathematics teacher can leave home without then arriving late at school. *(3 marks)*

5 A random sample of 200 university students reading Mathematics revealed 98 students with an IQ in excess of 130. A random sample of 200 university students reading English revealed 74 students with an IQ in excess of 130.

(a) Test, at the 5% level of significance, the claim that the proportion, p_M, of university students reading Mathematics with an IQ in excess of 130 is the same as the proportion, p_E, of university students reading English with an IQ in excess of 130. *(6 marks)*

(b) (i) Construct an approximate 99% confidence interval for $p_M - p_E$, giving the limits to four decimal places. *(2 marks)*

(ii) Hence state why a similar test to that in part (a), but with a 1% significance level, would produce a different conclusion from that obtained in part (a). *(2 marks)*

6 (a) The random variable X has unknown mean μ and unknown variance σ^2.

A random sample of size n, denoted by

$$X_1,\ X_2,\ \ldots\ ,\ X_n\ ,$$

has mean $\overline{X}$ and variance V, where

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad V = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2.$$

(i) Show that

$$\mathrm{E}(X_i^2) = \sigma^2 + \mu^2 \qquad \text{and} \qquad \mathrm{E}(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2.$$ *(2 marks)*

(ii) Hence show that $\dfrac{nV}{n-1}$ is an unbiased estimator of σ^2. *(3 marks)*

(b) Find an unbiased estimate of the population variance, σ^2, in the case where

$$\sum x = 150, \qquad \sum x^2 = 2700 \qquad \text{and} \qquad n = 10\,.$$ *(3 marks)*

END OF QUESTIONS