

General Certificate of Education  
January 2005  
Advanced Level Examination



**MATHEMATICS (SPECIFICATION A)**  
**Unit Statistics 4**

**MAS4/W**

Tuesday 1 February 2005 Morning Session

**In addition to this paper you will require:**

- an 8-page answer book;
- one sheet of graph paper for use in Question 4;
- 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.

**Advice**

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

---

Answer **all** questions.

---

1 In a poetry competition, six judges each rank the finalists 1 (lowest) to 7 (highest). The judges are **not** allowed to use tied ranks.

(a) The ranks given to the finalists by judges 1 and 2 are as follows.

Finalist	A	B	C	D	E	F	G
Judge 1	1	2	3	4	5	6	7
Judge 2	4	2	6	7	1	3	5

(i) Calculate the value of Spearman's rank correlation coefficient,  $r_s$ , for judges 1 and 2. (4 marks)

(ii) Comment on the value of  $r_s$  that you have obtained. (1 mark)

(b) (i) The value of  $r_s$  for judges 3 and 4 is found to be +1. What can be said about the rankings given by these two judges? (1 mark)

(ii) Write down rankings given by each of judges 5 and 6 which will result in a value for  $r_s$  of  $-1$ . (1 mark)

2 The table below shows the points for certain letters in Scrabble together with the percentage usage of these letters from a random sample of text.

Letter	A	G	M	W	K	X	Q
Points	1	2	3	4	5	8	10
Percentage usage	8.2	2.0	2.4	2.4	0.8	0.2	0.1

(a) Calculate the value of the product moment correlation coefficient between the points and the percentage usage. (5 marks)

(b) Assuming that these data are a random sample from a distribution with correlation coefficient  $\rho$ , investigate, at the 5% level of significance, the hypothesis that  $\rho < 0$ . (4 marks)

(c) Comment on the way that points are assigned to letters in Scrabble. (1 mark)

- 3 A tennis player claims to have a probability of 0.6 of producing a successful first serve. A random sample of 25 of the player's first serves produced 11 which were judged to be successful.

Stating hypotheses, investigate, at the 5% level of significance, whether this shows evidence that the success rate of the player's first serve is less good than claimed. *(6 marks)*

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

A psychologist randomly selects 8 rats. He gives each rat an amount,  $x$  hours, of training at exiting mazes from their centres. He then records the time,  $y$  seconds, for each rat to exit a test maze. The results are shown in the following table.

$x$	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0
$y$	62	74	51	49	28	35	28	26

- (a) Plot a scatter diagram of these data. *(2 marks)*
- (b) Find the equation of the least squares regression line of  $y$  on  $x$  and draw it on your scatter diagram. *(6 marks)*
- (c) (i) Hence estimate  $y$  when  $x = 2.5$ . *(1 mark)*
- (ii) Comment on the likely accuracy of your estimate. *(1 mark)*
- (d) Give **two** reasons why the model would be unsuitable to predict  $y$  for much larger values of  $x$ . *(2 marks)*
- 5 (a) A random sample of 100 shoppers at a department store in a certain town showed that 16 used the Park and Ride scheme.

Stating your hypotheses, test, at the 5% level of significance, the Local Authority's claim that 21 per cent of shoppers use the Park and Ride scheme. *(6 marks)*

- (b) A random sample of 100 shoppers at another department store in the same town showed that 19 used the Park and Ride Scheme.

Calculate an approximate 99% confidence interval for the difference between the proportion of shoppers at this department store using the Park and Ride scheme and the proportion of shoppers at the department store in part (a) using the Park and Ride scheme. Give the limits to three decimal places. *(5 marks)*

- 6 The proportion,  $p$ , of defective cathode-ray tubes from a production line can be estimated by taking a random sample from those produced.

The random variable  $X_1$  denotes the number of defective cathode-ray tubes in a random sample of size  $n_1$ .

(a) (i) Show that  $P_1$ , where  $P_1 = \frac{X_1}{n_1}$ , is an unbiased estimator of  $p$ . (2 marks)

(ii) Find the variance of  $P_1$ . (2 marks)

- (b) The random variable  $X_2$  denotes the number of defective cathode-ray tubes in a second random sample, of size  $n_2$ .

Given that  $P_2 = \frac{X_2}{n_2}$  and  $P = \frac{1}{3}(2P_1 + P_2)$ , show that:

(i)  $P$  is an unbiased estimator of  $p$ ; (2 marks)

(ii)  $\text{Var}(P) = \frac{p(1-p)}{9} \left( \frac{4}{n_1} + \frac{1}{n_2} \right)$ . (3 marks)

- (c) (i) Find the range of values of  $\frac{n_1}{n_2}$  for which

$$\text{Var}(P) < \text{Var}(P_1) \quad \text{and} \quad \text{Var}(P) < \text{Var}(P_2). \quad (3 \text{ marks})$$

- (ii) Given that  $n_1 = 300$  and  $n_2 = 100$ , state, with a reason, which of  $P_1$ ,  $P_2$  and  $P$  would be a best estimator of  $p$ . (2 marks)

**END OF QUESTIONS**